The London School of Economics and Political Science

Social Capital in Historical Perspective: A Principal Components Approach to International Measurement and its Economic Implications, 1870-2000

Marta Felis Rota

A thesis submitted to the Department of Economic History of the London School of Economics for the degree of Doctor of Philosophy, London, September 2008

ProQuest LLC
789 East Eisenhower Parkway
P.O. Box 1346
Ann Arbor, MI 48106-1346

Declaration

I certify that the thesis I have presented for examination for the MPhil/PhD degree of the London School of Economics and Political Science is solely my own work other than where I have clearly indicated that it is the work of others (in which case the extent of any work carried out jointly by me and any other person is clearly identified in it).

Abstract

The thesis proposes measuring social capital by means of Principal Components Analysis (PCA) and presents international social capital estimates for the late-nineteenth century in the form of a Social Development Index (SDI) for 1870 and 1890. The analysis is based on a nineteenth-century international database containing a wide range of socio-economic variables. These indicators are compared to mid-twentieth century social indicators, facilitating the study of the evolution of social capital in the nineteenth and twentieth centuries. One of the conclusions that arise from the PhD thesis is the persistence of social indicators in the medium term. In the very long run, a significant decline in the relative position of the Western European countries and the United States is found.

The new series are then used as empirical evidence to show that comparative social evolution trajectories can improve economic performance predictions. In fact, the results show that the relationship between social development and per capita income already existed in the late-nineteenth century. This relationship upholds after controlling for foreign trade volume and structure, urbanization, education, quality of institutions, political stability, government expenditure, population growth, and climate.

Finally, the thesis presents an application of the proposed SDI into an Economic Geography framework. In particular, it shows how a better definition of the formal and informal institutional setup can help testing economic geography hypothesis on international trade. In order to do this, indicators of access to markets are calculated for 1994 and other benchmark years going back to 1962 making use of bilateral trade data. Then, the proposed alternative socio-institutional measures illustrate how the importance of access to markets is not unconditional on the institutional setup. Poor government effectiveness undermines the positive benefits of geographical location. Only when a country reaches a certain threshold of institutional quality can it effectively benefit from location.

Acknowledgements

I would like to thank everyone who gave support to this project throughout the years, either directly or indirectly, by reading the material, attending to seminars, giving words of encouragement and/or criticism, and very specially to my doctoral supervisor, Professor Nicholas Crafts.

This project was funded by the London School of Economics and Political Science (LSE) and the Economic and Social Research Council (ESRC).

I am eternally indebted to all of them.

TABLE OF CONTENTS

LIST OF TABLES, GRAPHS AND MAPS

Chapter 1

A NEW PERSPECTIVE ON SOCIAL CAPITAL

1. I. INTRODUCTION

Social capital is important. The classical thinkers were already giving importance to living in a community rather than in isolation. Giambattista Vico illustrates the power of society in the following passage: '[society] makes national defence, commerce, and politics, and thereby causes the strength, the wealth, and the wisdom of the republics; (...) society thus causes the civil happiness to emerge. (...) [Men] who are entirely occupied by the pursuit of their private utility are transformed into a civil order which permits men to live in human society' (Vico in *Scienza Nuova*, paragraphs 132-33, from Hirschman, 1977:17). Such is the power of society; this device makes man work for the common good while they paradoxically pursue their own benefit. In Montesquieu's words, 'it turns out that everyone contributes to the general welfare while thinking that he works for his own interests' (Montesquieu, *Esprit des Lois*, Book III, Ch VII, as cited in Hirschman, 1977: 10).

Indeed, social arrangements play an important role in security, knowledge and education, and the protection of property rights. Society can generate the adequate atmosphere for economic prosperity, as the development of countries who have historically promoted these values shows. Western European countries have been careful in ensuring security across history, from feudalism to nowadays. Exemplifying further the importance of social arrangements, one of the key elements of the recent growth experience in South East Asia has been the role of the state in providing education. Finally, the United States of America are a clear example of early protection of the property rights, which, among other things, boosted the rate of innovation with the help of the patent laws. John Stuart Mill emphasised this positive role of the social arrangements, as well has their role in generating positive attitudes toward work and saving (Mill, 1909:108-110).

John Stuart Mill emphasised also the benefits from trade (Mill, 1909:581). However, 'the spirit of liberty appears (...) to depend chiefly upon two circumstances: first, the condition of the people relative to the distribution of property, and the means of subsistence: secondly, the facility with which the several members of society are enabled to associate and to act in concert with one another' (Lehmann, 1960: 330-1). Notice that Lehmann gives as much importance to the distribution of wealth and availability of resources as to people interactions; so the relevance of the latter is not a

new idea. Nevertheless, it has been recently recovered, reshaped, redefined, and put at the steak for discussion.

1. II. SOCIAL CAPITAL: DEFINITION

The importance of the structure of society and of the interactions among its agents has been recognised as a relevant factor in the performance of economies. The social links between citizens can be encompassed under the concept of 'social capital' (DiPasquale and Glaeser, 1999: 355). In the words of the political scientist Robert Putnam, social capital 'refers to features of social organisation, such as trust, norms, and networks, that can improve the efficiency of society by facilitating co-ordinated actions' (Putnam, 1993: 167). Margaret Keck and Kathryn Sikkink define networks as '"forms of organization characterized by voluntary, reciprocal and horizontal patterns of communication and exchange" [Keck and Sikkink, 1998:8]. These networks may include not only conventional NGOs, but also local social movements, foundations, the media, churches, trade unions, consumer organizations, intellectuals, parts of regional and international intergovernmental organizations, and parts of the executive and/or parliamentary branches of governments' (Brecher, Costello, and Smith, 2000:83). In Woolcock's (1998) opinion, however, these features are not social capital, but rather the consequences thereof. He adds that the concept can be read in several ways.

Woolcock argues that there are four dimensions of social capital: the size and scope of horizontal associations; social integration, the nature of social ties within communities; the relationship between civil society and the state; and, the quality of governing institutions. However, he agrees that it should be collapsed and understood as a single variable. Fine cites a statement made by Woolcock on the World Bank's email discussion site on social capital:

> 'Several critics, not without justification, have voiced their concern that collapsing an entire discipline into a single variable (especially one with such economic overtones) is a travesty, but there are others who are pleased that mainstream sociological ideas are finally being given their due at the highest levels'(Fine, 2001:139).

As a way of summarising the various proposed definitions, social capital can be understood in the abstract as the quality or health of the interactions within civil society.

Despite its recent re-definition and re-contextualisation, the concept of social capital is nothing new. Putman acknowledges the fact that the concept was described well before his influential 1993 book. 'The term *social capital* turns out to have been independently invented at least six times over the twentieth century, each time to call attention to the ways in which our lives are made more productive by social ties' (Putnam 2000:19). However, one has to bear in mind that some of the claims on social capital are retrospective claims, which affirm that 'writers (...) really were talking about social capital when they thought they were writing about something else' (Harriss, 2001:75).

The 'six times' Putnam acknowledges are the following, starting by L. J. Hanifan in 1916, who studied the importance of the community involvement for the well-functioning of schools. A forty years gap separates Hanifan from the book co-authored by John Seely, Alexander Sim, and Elizabeth Loosley in 1956, *Crestwood Heights: A Study of the Culture of Suburban Life.* Keeping on the line of the study of urban societies and neighbourliness, Jane Jacobs presented her work on 1961. It was not until the mid-seventies when Glenn Loury rediscovered social capital once more, writing about the determinants of income differences between members of different ethnic groups (Loury, 1977). Loury's discovery attracted the attention of the most influential of the six, James Coleman, who defines social capital according to Loury's vision as 'the set of resources that inhere in family relations and in community social organisation and that are useful for the cognitive or social development of a child or young person' (Coleman, 1990: 300). The sixth, according to Putnam, was the Marxist theorist Pierre Bordieu who underlined the relevance of social networks in the 1980s. In fact, as noticed by Harriss (2001), Karl Marx already envisaged capital as a social relation. Thus, the twentieth century term 'social capital' is conceptually not that far from Marx's broad view of 'capital' itself.

Due to the relatively recent usage of the term 'social capital', it might be misunderstood or mixed with other close concepts. To avoid confusion, it is important to make clear the difference between social capital and institutions. Personal interrelations or interactions are the action while institutions (regulations and conventions) are the framework in which the actions are embedded. At the same time, interactions contain or make use of the stock of social capital available in a society. In a few words;

interactions making use of social capital may take place within the institutional framework.

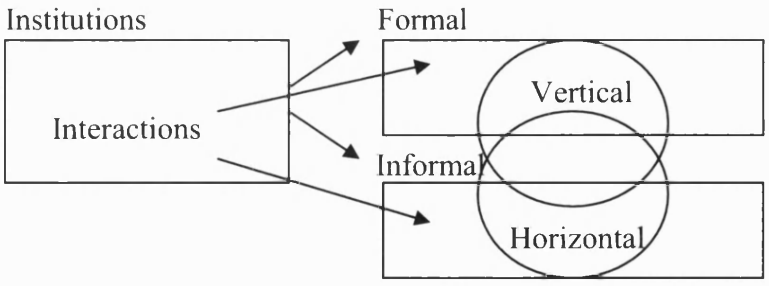
It is convenient to underline the double dimensionality behind the combination of the two concepts above: institutions and interactions.

1) On the one hand, institutions might be either formal or informal. Formal institutions are the ones established and governed by law. Informal institutions are the networks or groups of people which are spontaneously generated and still host action on a regular basis. These may or may not have written rules and, if this is the case, these would be voluntarily joined.

2) On the other hand, interactions may occur at and between different hierarchical levels. Vertical interrelations or interactions are typically at the basis of formal institutions, since superior hierarchies have the power and acquired rights to dictate the law, which is a necessary requirement to create a formal institution. Horizontal organisations are typically informal; they are not forced upon individuals by law but joined spontaneously and voluntarily instead.

See figure 1.1 for a schematic explanation.

Figure 1.1 - Double overlapping dimensionality of institutions and interactions



Institutions can be either formal or informal. Interactions can be either vertical or horizontal. Social capital typically lies within the informal and horizontal frameworks.

This double dimensionality both defines and classifies the space of events. Social capital typically finds its place within the horizontal interactions and informal type of

institutions. To avoid restricting myself to the recently created 'social capital' branding, I will also refer to 'social capability' or 'social development' as broader, less encapsulated terms for the same idea.

1. III. THE IMPORTANCE OF SOCIAL ARRANGEMENTS

The next pages summarise a selection of cases in which social ties or civic virtues have contributed towards a positive outcome in several ways. Section III.1 relates social capital to some empirical studies in economics. Section III.2 provides some examples of the importance of social ties in the distant past. Historical evidence has started to be slowly incorporated into economic theory and this indeed could now be the way to go.

1. III.1. SOCIAL CAPITAL AND ECONOMIC GROWTH

The relevance of social capital reflects the importance of people. Economic organisations are managed by people, and it is through their interactions that economy works. Among all the schools of thought trying to explain development and economic growth, one stream presents some interesting interactions between sociology and long-run economic development; particularly focusing on social capabilities, as identified by Abramovitz (1986:387-8).

There is some empirical evidence that social capital has had significant influence on economic growth. Studies conducted by Easterly and Levine (1997), and Temple and Johnson (1998) aimed at determining the macroeconomic consequences of social arrangements. They indicate that some disappointing policy outcomes have their roots in the nature of societies. La Porta *et al.* (1997) find some effect of social capability on growth for the period 1970-93; Knack and Keefer (1997) focus on the period 1980-92, and find a stronger link: If the level of trust increases by 10 percent, growth is, on average, 0.8 percent higher per year. This is a very considerable effect. Therefore, it stands clear from the existing research that social capital has some economic consequences, and it makes the difference as far as economic growth is concerned.

Following on the potential lines of influence of social capability, its effects on financial development had been empirically tested for Italy by Guiso, Sapienza, and Zingales. Using microeconomic data on households and firms, they prove that 'in areas of the

country with high levels of social trust, households invest less in cash and more in stock, use more checks, have higher access to institutional credit, and make less use of informal credit. (...) Firms also have more access to credit and are more likely to have multiple shareholders',(Guiso, Sapienza, and Zingales, 2000: 0). Their empirics also suggest that trust seems to be less important in areas with poor legal systems and more educated people. Therefore, 'social capital is likely to be very important in explaining the success (or lack thereof) of developing countries' (Guiso, Sapienza, and Zingales, 2000: 34).

Inquiring about the mechanism behind the economic effects of social capital, one important line of research is its role in making institutions work well, through a reduction of transaction costs; '(that is, the costs of monitoring and enforcing agreements), and thus in enabling agents more efficiently to surmount problems of opportunism and shirking' (Putnam, 93:166). Harriss recalls Coleman's example of traders in a Cairo market 'who share information about customers – he shows how the reciprocity and trust which may be an aspect of social relations are of value because they help to reduce many of the costs of transaction, through the communication of information and the kinds of insurance that are created in social networks' (Harriss, 2001:5). This could potentially explain why institutions work better in some places than in others. Institutions, as technology, may be replicated, but yet they achieve different levels of efficiency in different places. This depends among other factors upon the social capability of the society in question. Thus, one of the products of social capital is that it helps institutions work well.

In order to capture social capability in the empirics, we need first to understand how it is generated. A lot has been done about its effects, but very little into its causes. It is important to understand what influences social capital itself and, hence, much more investigation needs to be done into the causes that give rise to different levels of social capital. What can the differences in social capital be attributed to? As showed in *Making Democracy Work*, '"civic involvement 1900s" predicts "socio-economic development 1970s" more strongly than does "socio-economic development 1900s"'(Harriss, 2001:21). 'A region's chances of achieving socio-economic development during this century have depended less on its initial socio-economic endowments than on its civic endowments. In so far as we can judge (...) the contemporary correlation between civics and economics reflects primarily the impact of civics on economics, not the reverse'

(Putnam, 1993: 157). The direction of the causality is hard to prove, but Putnam's research is a good piece of evidence pointing at the existing impact of society into economics.

The process itself of the generation of social capital could be explored further in a posterior stage *via* modelling. Glaeser, Laibson, and Sacerdote (2002) present a simple model of optimal investment that explains individual's decision to invest in the creation of social capital. Also, a general equilibrium model is presented by Zak and Knack (2001), where agents have the opportunity to either trust others agents or to invest in verifying other people's actions. So, there are some initiatives focusing on explaining the mechanism generating social capital and enquiring about causality. This thesis, however, focuses on the measurement issues and impact on economic development at the macro level instead.

Let us take the growth regressions literature as the initial framework, in which social capability still has to find its place. The comprehensive paper by Bleaney and Nishiyama (2002) constitutes a compendium of the most significant advances in the 1990s growth literature that try to incorporate new variables to the growth regressions. They build up a new benchmark model, which they call an 'encompassing model', that arises from a selection of the most significant variables coming from three different previous non-nested models: Barro (1997), —an updated version of his influential work in 1991—, Easterly and Levine (1997), and Sachs and Warner (1997). They also test the non-nestedness of these three, which is confirmed.

The variables that Bleaney and Nishiyama include in their encompassing model are referring to the initial per capita income (taking 1965 as the year of reference), openness, life expectancy, schooling, institutional quality, democracy, government savings, primary products exports, climate and active population growth. This model passed several robustness tests, including the incorporation of regional dummy variables. So, the influence of these variables on economic performance has already been tested and given approval, but most of the mechanisms that lead to interactions and intermediate links are still unexplored. Here is where social capital might have a role and it is therefore worth exploring it.

1. III. 2. SOCIAL CAPITAL IN THE DISTAN PAST: SOME HISTORICAL EXAMPLES AND NATURAL EXPERIMENTS

Is it worthwhile making the effort to try to capture such an abstract variable? In order to answer this question, the following paragraphs expose a series of historical examples and natural experiments which illustrate how social capital may matter.

The next historical example shows how informal links between people of different villages could have been the key to the winning warfare strategy leading to the foundation of the Zulu state. The origin of the Zulu state by the beginning of the nineteenth century in south-eastern Africa shows how it is possible to detach by other means than technological. 'Among the dozens of Zulu chiefdoms, the Mtetwa chiefdom enjoyed no advantage whatsoever of technology, writing, or germs over the other chiefdoms, which it nevertheless succeeded in defeating. Its advantage laid solely in the spheres of government and ideology' (Diamond, 1997:292). The chief of the Mtetwa, called Dingiswayo,

> 'developed a superior centralised military organisation by drafting young men from all villages and grouping them into regiments by age rather than by their village. He also developed superior centralised political organisation by abstaining from slaughter as he conquered other chiefdoms, leaving the conquered chief's family intact, and limiting himself to replacing the conquered chief himself with a relative willing to cooperate with Dingiswayo' (Diamond, 1997: 290).

Whether Dingiswayo mixed people from different villages on purpose or just as part of a military logistic tactic, his advantage laid specifically in creating links among people belonging to different villages, building in this way mutual trust and cooperation instead of promoting rivalry. It is a pity that Diamond skips to underline this point in his argumentation, arguing just superiority in political centralisation.

'Among large tribes, those with stronger big-men and hence greater centralisation tend to have an advantage over those with less centralisation' (Diamond, 1997:288). This advantage consists not only of political and military superiority, but also of better economic organisation and possible higher levels of social capability. To give an example of the latter, 'water control systems (...) appear to have been associated with

centralised political organisation in some (...) areas of the world, including the Indus Valley of the Indian subcontinent, the Yellow and Yangtze Valleys of China, the Maya lowlands of Mesoamerica, and the coastal desert of Peru'. Given the fact that both the Maya lowlands and highlands were part of the same civilisation, why did the former develop water control systems while the latter did not? 'Detailed archaeological studies have shown that complex irrigation systems did not accompany the rise of centralised bureaucracies but followed after a considerable lag. That is, political centralisation arose for some other reason and then permitted construction of complex irrigation systems' (Diamond, 1997:23). This makes us wonder whether political centralisation pushed society towards higher levels of trust and social capital, which endowed society with a greater ease for coordinated actions.

Here is another example of how social capital is required to establish the foundations to make institutions work. We know from Clifford Geertz (1962) reports from Java peasants that 'rotating credit associations are [were] often found in conjunction with cooperatives and other forms of mutual aid and solidarity. In part, this is [was] because all these forms of voluntary cooperation are [were] fed by the same underlying stock of social capital' (Putnam, 1993:169).

On the contrary, a fragmented society may deter important regions from significant economic growth. 'New Guinea has by far the highest concentration of languages in the world: 1,000 out of the world's 6,000 languages, crammed into an area only slightly larger than that of Texas, and divided into dozens of language families and isolated languages as different from each other as English is from Chinese'. 'Difficulties of terrain, combined with the state of intermittent warfare that characterised relations between New Guinea bands or villages, account for traditional New Guinea's linguistic, cultural, and political fragmentation' (Diamond, 1997: 306). We could argue a similar situation in twentieth-century Europe: 'Difficulties of terrain, combined with the state of intermittent warfare'. Though fragmentation is much lower than in New Guinea we can still think that this could have contributed to the dominance of larger unified nations such as the United States in the twentieth century, or China in the twenty-first century, according to prospective growth rates.

Recent studies have moved towards natural experiments, such as Costa and Kahn (2007) and Fisman and Miguel (2006). The first finds that friendship and other links

such as kinship or place of origin favoured survival of prisoners of war captured in 1863 during the American Civil War. The second makes use of diplomatic immunity in order to test for civic virtues, beyond any legal enforcement. Fisman and Miguel find out that diplomats coming from high corruption countries have a stronger tendency to violate parking norms when visiting New York City. These are two examples of how social ties or identities do actually influence behaviour.

1. IV. SOCIAL CAPITAL MEASUREMENT

> 'It is much easier to describe and be precise about the formal rules that societies devise than describe and be precise about the informal ways by which human beings have structured human interaction. But although they defy (...) neat specification and it is extremely difficult to develop unambiguous tests of their significance, they are important' (North, 1990:36).

In this passage, Douglass North points out the importance of informal organisations and *modus operandi* of human activity. However, he does not take this statement further, he does not explore the phenomenon in more detail.

The duty of the social scientist is to analyse the social phenomena. No matter how imperfect the measurement of social phenomena is, these should be subject to non-quantitative as well as quantitative analysis. So, even if the quantitative analysis procedure is far from optimal, it might still reveal some interesting stylised facts or links that invite us to look for further detailed evidence and improve the procedure itself. It is the case, indeed of the social arrangements. In other words, taking North more seriously than he did himself could constitute a contribution to the theory of institutions and growth.

Taking North's forceful statement together with the argumentation above, we can start by concluding that there is a strong enough theoretical basis which encourages us to pursue further quantitative study of social capital. There are some recent studies relating social capital to economic development. Still, the study of social capital from a long run perspective is just starting and deserves more effort, especially from a quantitative perspective. Researchers need to bear in mind that definition and measurement walk hand by hand. In order to be able to carry on some statistical study on social capital,

there is not enough by stating a definition. It is necessary to obtain a quantitative measure.

The abstract character of the notion makes social capital very difficult to quantify in the empirics. Social capital is neither an implement nor can it be found in isolated individuals. Rather, it is an intangible collective good. Unlike other qualitative variables, social capital is not trivially dividable in different conceptual components, nor is it possible to be transformed into a yes/no variable. For instance, democracy can be set as an attribute of a certain political regime. It is a qualitative attribute, and can therefore be transformed into a 0-1 variable to become a ready-to-use variable for statistical analysis. On the contrary, all societies possess social capital to a certain extent. Some measurement proposals have been made with the final goal of linking social capital — or another closely related variables, which could act as a proxy—, to economic performance. However, no proposal enjoys general acceptance amongst social scientists.

The first related cross-section quantitative study was undertaken in the 1960's by Adelman and Morris, (see 1965 and 1967 publications). They started from the idea that some societies may be inherently more suited to entrepreneurship and economic development. Societies evolve in a variety of dimensions, such as family relations, the extent of communications, the importance of the middle class, and social mobility. Departing from this idea, they used a series of socio-economic variables in a cross-country historical investigation in order to explain different patterns of development. In fact, moving to more recent studies, Temple and Johnson (1998) have proposed the use of the social development index elaborated by Adelman and Morris as a proxy for social capital. Other empirical work has focused on international differences in the level of trust, and of civic co-operation. The measure of social capital that Putnam uses in his book *Bowling Alone* 'is constructed – like that of "civic community" in *Making Democracy Work* [his previous book]—by combining a number of indices (of "community organizational life", "engagement in public affaires", "community volunteerism", "informal socializing" and of "social trust") which are themselves highly intercorrelated' (Harriss, 2001:55). Papers like that of Alesina and La Ferrara (2002) focus more narrowly on the level of trust for the United States in the period 1974-1994. The level of trust is established by the General Social Surveys (Davis and Smith, 1994), which asked to the respondents whether they believed or not that most people can be

trusted. Other alternatives are experiments where civic virtues are tested by, for example, "losing" a wallet with fifty dollars in different cities and counting the percentage of wallets being returned (Knack and Keefer, 1997). In their paper 'Are Homeowners Better Citizens?', DiPasquale and Glaeser (1999) point at the importance of being a homeowner for the involvement in the social initiatives aiming at the well-being of the local community. Thus, it shows clearly that many proposals are available, yet none enjoys general acceptance.

Some recent work is starting to pay attention to social capital or related concepts from a long run perspective. On the one hand, Putnam (2000) proposes to measure social capital in the past with things like newspaper readership or belonging to voluntary associations as a way to measure the level of interaction in a society. However, this is a study of the United States only and it is difficult to transfer to the distant past. On the other hand, Tabellini (2007) achieves a more global perspective by instrumenting culture with distant past institutions. He does not refer to social capital but to culture instead, though his definition of culture is very close to social capital. So there are some recent efforts to integrate social capital or related concepts like culture within the economic literature and bring some historical perspective at the same time. However, there is no homogeneous and/or consensuated measure or indicator. This project proposes an attempt towards international measurement of social capital at times where the current indicators based on surveys are not available.

1. V. A NEW PERSPECTIVE

Going back to the starting point, the Adelman and Morris work could shed some light on the questions in the agenda. Now that we can look at their analysis with several decades of distance, it appears to have not only the historical insights the authors were presenting but also more predictive power than they suspected. Their indicators for the 1960s turned out to be *apparently* too pessimistic for Latin America and too optimistic for some Far Eastern countries, and so they moved along omitting some indicators (Adelman and Morris, 1967). At that time, they did not know the importance of their findings. Their method could have helped forecasting subsequent growth better than any other contemporary attempt. Temple and Johnson (1998) tell us that the socio-economic index constructed by Adelman and Morris for 1960 'could have helped researchers

make much better forecasts of long-run growth rates'. So, it looks promising to bring their research further.

Adelman and Morris' research methodology can be applied to social capital measurement because it establishes a procedure for generating a social index. They created a social index for 1960, and they also used a very similar approach in their studies on comparative economic development in the nineteenth century, but much more can be done. Their nineteenth century database, which was constructed and initially explored with the intention of explaining economic development has possibilities for being useful in the creation of an international social index for the nineteenth century.

1. V.1.WIDENING OF THE TOOLKIT: DUAL QUANTITATIVE METHODOLOGY

One of the objectives is to present the quantitative results in a readable way for historians. As Nicoli Nattrass once wrote,

> 'historians and sociologists tend not to read specialist economic journal and when they encounter arguments with economic content are alienated by the terminology and underlying methodology. For their part many economists have contributed to this by regarding sociology and history as unscientific and largely irrelevant to economic issues, (...)'. (Nattrass, 1991: 655).

The use of quantitative methods in economic history was popularised around half a century ago. I strongly agree with Mitchell's statement: 'Statistics are used nowadays (...) as a major raw material of much economic history, especially of economic growth. (...) The subject is by its very nature concerned with quantities, (...) though no sensible historian would ever claim that statistics can tell the whole story' (Mitchell, 1998: vii). The inverse is also true; economic historians must utilise historical reasoning, but there is a certain reluctance to accept plain grammatical arguments from the most economic-theory oriented focuses. I would like to give to my thesis a reconciling approach.

As far as economists are concerned, the methodology with which economic growth research has generally been addressed has been narrow-ranged. The main weaknesses

have typically been the use of a self-restrictive methodology and the lack of historical perspective.

Empirical macroeconomists' methodology can be typically encapsulated in one of two types: model calibration or econometric regressions. Real Business Cycle literature and contemporary monetarist economists use model calibration. The representative paper by Barro (1991), and the more recent Bleaney and Nishiyama (2002) are examples of the empirical growth regressions' literature, related to our subject. Growth regressions consist on running multiple regressions of growth rates on a list of variables in order to identify which ones are more correlated with economic growth. This generalised technique is able to identify which variables move together, provided the researcher did not omit any important variable in the multiple regression. Infinitely many variables cannot be included, since the data sample has a limited amount of degrees of freedom. Therefore, the researcher needs to choose between the trade off of including extra variables and allowing for more degrees of freedom. An important limitation of the regression analysis is that a specific functional form needs to be specified. In most of the growth regressions, the functional form of the regression remains restricted to a merely linear or at most quadratic form. Therefore, the growth regressions technique is helpful and broadly utilised, but used alone could lead to a narrow view of the interrelations amongst variables.

I would like to propose the re-introduction of a third empirical toolbox: the principal components analysis; which was first used in development studies by Adelman and Morris (1965). This approach is not meant to substitute the traditional approaches, but is intended to provide a complementary perspective for the better understanding of the underlying mechanisms and interrelations amongst variables. I defend it as especially appropriate for the historical understanding of complex socio-economic relations.

The re-introduction of the principal components analysis, in addition to the use of the traditional econometric tools, endows this project with a dual quantitative methodology. In this way, I proceed in two steps, the first step being the principal components analysis and the second one being the econometric regression analysis. The dual methodology undoubtedly enriches the project beyond the scope of any one-dimensional quantitative methodology.

I would like to remark that the principal components analysis' technique (PCA) is not intended for hypothesis testing. Instead, it is designed for *building* hypothesis, since it allows the researcher a lack of a priori model specification in the first instance. Therefore, in this first quantitative stage we can enjoy doing without the self-restrictive *a priori*, which are the basis of econometric estimation and hypothesis testing. Instead, a statistical software deals with a set of interrelated variables in which everything affects everything, and extracts the relevant hidden dimensions. The job of the economic historian beyond the statistician is to provide the initial set of variables and be able to interpret the results.

In our case, econometric regression analysis will focus on a selection of variables arising from a studied combination of both the theoretical and historical survey, and the results obtained from the principal component analysis. Alike in the principal component analysis, equations in regression analysis are not produced by the software programme, but are part of the input; equations are derived from theory. New hypotheses are more than likely to appear while processing the data and after interpreting the first quantitative results.

Due to the complexity of the calculations with a large dataset, we want to make use of computer support. A number of statistical packages have been designed for the practical implementation of these techniques. These are able to process the numerical calculations as well as producing self-explanatory graphs.

Finally, I would like to add a few words on the numerical representation of the chosen variables. Some variables are very easy to represent numerically because they are numerical in nature, like income or annual percentage of growth of the economic activity. Others are qualitative in nature, which is frequently the case in sociology and history. These can be represented by indices (weighted average of a set of defining features that can be quantified), for instance institutional quality and social capital; or dummy variables (0-1 variables, where a given observation takes the value 0 if a certain statement is false and 1 if it is true), such as democracy. We make use of dummy variables in order to numerically represent qualitative variables that can be represented with a limited number of categories, such as political regime, continent or climate. This technical device allows capturing different effects of social capital on economic growth depending on climate, for instance. Another example: If we introduce a dummy variable

indicating 'wartime', we will be able to obtain a completely new set of coefficients under this circumstance.

1. V.2. WIDENING THE HISTORICAL HORIZON

As far as the historical perspective is concerned, the generalised use of the Penn World tables, —constructed initially by Heston and Summers (1991)—, has limited the time span of most of the studies to a maximum of forty to fifty years. When we think of a long run empirical work, the homogeneity and accessibility of these tables do not compensate for the fact that they do not start until the post Second World War period.

The database constructed by Adelman and Morris contains a wide range of socio-economic variables, with data from the nineteenth century. As this database was constructed forty years ago, it is possible to update it with more recently available material. The immediate consequence would be more than doubling the conventional time span for this type of study. Together with the contemporary studies in social capital, we could encompass more than one hundred years of data (with gaps), hence achieving a much broader historical perspective.

1. V.3. HOW IS THE NEW PERSPECTIVE ACHIEVED?

The first task will consist of reconstructing and updating the database from Adelman and Morris. After digitalising it, the possibility of compiling data that has recently become available from posterior historical country studies is considered. The new database will allow the generation of a new social index for the nineteenth century and testing the theory about the economic impact of social capabilities in a way that has never been done. First, we face the question of whether growth trajectories support the hypothesis relating economic development to social capabilities in the past. Then, a final part of the thesis shows that a social index for the past can be useful in economic theory explorations. In fact, the index proves useful in helping to test economic theory hypothesis in historical perspective.

It is very important to understand that we are in front of a global project. It is not a case study or a few isolated case studies. Case studies are only utilised for the sake of argumentation, illustration, and inspiration of the general laws put to the test. As

Diamond put it, 'prediction in history, as in other historical sciences, is most feasible on large spatial scales and over long times, when the unique features of millions of small-scale brief events become averaged out' (Diamond, 1997: 424). Given the global theoretical nature of the hypothesis, this seems more appropriate. The intention of the thesis is to provide a cross-country global overview at several time cuts.

This exploration aims at instrumenting the quantification of social arrangements. Amongst the interesting conclusions that may arise, could be the finding of new historical considerations entering the process. This would partly be accomplished thanks to the enlargement of the time period concerned, and partly thanks to the generation of new empirical results supporting the theory. Both dimensions potentially lead to the discovery of new underlying facts about the way the economy works and has worked in the past. The final result is the enlargement of the body of knowledge about relationships among variables.

REFERENCES[1]

Abramovitz, Moses, 1986, 'Catching Up, Forging Ahead, and Falling Behind', *The Journal of Economic History*, 46 (2): 385-406.

Adelman, Irma, and Cynthia T. Morris, 1965, 'A Factor Analysis of the Interrelationship Between Social and Political Variables and Per Capita Gross National Product', *The Quarterly Journal of Economics*, 79 (4): 555-578.

———, 1967, *Society, Politics, and Economic Development: A Quantitative Approach*, John Hopkins Press (Baltimore).

Alesina, Alberto and Eliana La Ferrara, 2002, 'Who Trusts Others?', *Journal of Public Economics*, 85 (2): 207-234.

Banfield, Edward Christie, 1958, *The Moral Basis of a Backward Society,* The Free Press (Chicago).

Barro, Robert J., 1991, 'Economic Growth in a Cross Section of Countries', *Quarterly Journal of Economics*, 106 (2): 407-443.

———, 1997, 'Determinants of Economic Growth: A Cross-Country Empirical Study', MIT Press (Cambridge, Massachusetts).

Bleaney, Michael and Akira Nishiyama, 2002, 'Explaining Growth: A Contest Between Models', *Journal of Economic Growth*, 7: 43-56.

Bordieu, Pierre, 1983, 'Forms of Capital', in John. G Richardson, ed., *Handbook of Theory and Research for the Sociology of Education*, Greenwood Press, (New York): 241-258.

Brecher, Jeremy, Tim Costello and Brendan Smith, 2000, *Globalization from Below: The Power of Solidarity,* South End Press (Cambridge, Massachusetts).

Coleman, James S., 1990, *Foundations of Social Theory,* Harvard University Press (Cambridge, Massachusetts).

Costa, Dora, and Kahn, Matthew E., 2007, 'Surviving Andersonville: The Benefits of Social Networks in POW Camps', *American Economic Review*, 97 (4): 1467-1487.

Davis, J. A. and T. W. Smith, 1994, *General Social Surverys, 1972-1994*, produced by the National Opinion Research Center (Chicago), distributed by The Roper Center for Public Opinion Research (Storrs, Connecticut).

[1] Due to the fact that the thesis has several well differentiated parts, it seems more appropriate to list bibliographical references at the end of each chapter rather than a unified bibliography at the end.

Diamond, Jared, 1997, Guns, Germs, and Steel: A Short History of Everybody for the Last 13,000 Years, Chatto & Windus (Great Britain).

DiPasquale, Denise, and Edward L. Glaeser, 1999, 'Incentives and Social Capital: Are Homeowners Better Citizens?' *Journal of Urban Economics,* 45: 354-384.

Easterly, William, and Ross Levine, 1997, 'Africa's Growth Tragedy: Policies and Ethnic Divisions', *Quarterly Journal of Economics*, 112 (4): 1203-50.

Fine, Ben, 2001, *Social Capital versus Social Theory: Political Economy and Social Science at the Turn of the Millennium*, Routledge (London and New York).

Fisman, Ray and Edward Miguel, 2006, 'Cultures of Corruption: Evidence from Diplomatic Parking Tickets', http://elsa.berkeley.edu/users/emiguel/miguel_parking.pdf (accessed 13/09/2007).

Geertz, Clifford, 1962, 'The Rotating Credit Association: A "Middle Rung" in Development', *Economic Development and Cultural Change* 10: 241-263.

Glaeser, Edward, David Laibson, and Bruce Sacerdote, 2002, 'An Economic Approach to Social Capital', *Economic Journal,* 112(483):437–458.

Guiso, Luigi, Paola Sapienza, and Luigi Zingales, 2000, 'The Role of Social Capital in Financial Development', *NBER Working Paper Series,* 7563.

Hanifan, Lyda Judson, 1916, 'The Rural School Community Center', *Annals of the American Academy of Political and Social Science,* 67: 130-138.

Harriss, John, 2001, *Depoliticizing Development: The World Bank and Social Capital*, Leftword (New Delhi).

Heston, Alan, and Robert Summers, 1991,'The Penn World Table: An Expanded Set of International Comparisons, 1950-1988', *The Quarterly Journal of Economics*, 106 (2): 327-368.

Hirschman, Albert O., 1977, *The Passions and the Interests: Political Arguments for Capitalism before Its Triumph*, Princeton University Press, Princeton (New Jersey).

Jacobs, Jane, 1961, *The Death and Life of Great American Cities*, Random House (New York).

Keck, Margaret and Kathryn Sikkink, 1998, *Activists Beyond Borders: Advocacy Networks in International Politics*, Cornell University Press (Ithaca, New York).

Knack, Stephen, and Philip Keefer, 1997, 'Does Social Capital Have an Economic Payoff? A Cross-Country Investigation', *The Quarterly Journal of Economics*, 112(4): 1251-1288.

La Porta, Rafael, Florencio Lopez-de-Silanes, Andrei Shleifer, and Robert W. Vishny, 1997, 'Trust in Large Organisations', *American Economic Review Papers and Proceedings*, 87: 333-338.

Lehmann, William C., 1960, *John Millar of Glasgow, 1735-1801*, Cambridge University Press (Cambridge).

Loury, Glenn, 1977, 'A Dynamic Theory of Racial Income Differences', in P. A. Wallace and A. LeMund, ed., *Women, Minorities, and Employment Discrimination*, Lexington Books (Lexington, Massachusetts): 153-188.

Marx, Karl, 1867, *Das Kapital: Kritik der Politischen Oekonomie*, Friedrich Engels, ed., (Hamburg). (English translation: *Capital: A Critique of Political Economy*, vol. I, *The Process of Production of Capital*, Penguin, 1976).

Mitchell, Brian Redman, 1998, *International Historical Statistics: Europe 1750-1993*, Fourth edition, Stockton Press and MacMillan Reference Ltd, (New York and Basingstoke respectively).

Mill, John Stuart, 1909, *Principles of Political Economy with Some of Their Applications to Social Philosophy*. This edition: 1961, W. J. Ashley, ed., (New York).

Montesquieu, Charles de Secondat, 1872, *Esprit des Lois*, (Paris).

Morris, Cynthia Taft and Irma Adelman, 1988, *Comparative Patterns of Economic Development, 180-1914*, The John Hopkins University Press (Baltimore and London).

Nattrass, Nicoli, 1991, 'Controversies about Capitalism and Apartheid in South Africa: An Economic Perspective', *Journal of Sourthern African Studies*, 17 (4): 654-677.

North, Douglass Cecil, 1990, *Institutions, Institutional Change and Economic Performance*, Cambridge University Press (Cambridge, UK).

Putnam, Robert D., 1993, *Making Democracy Work: Civic Traditions in Modern Italy*, Princeton University Press, (Princeton, New Jersey).

Putnam, Robert D., 2000, *Bowling Alone: The Collapse and Revival of American Community*, Simon & Schuster (New York).

Sachs, Jeffrey D., and A. Warner, 1997, 'Sources of Slow Growth in African Economies', *Journal of African Economies*, 6: 335-376.

Seeley, John R., Alexander R. Sim, and Elizabeth W. Loosley, 1956, *Crestwood Heights: A Study of the Culture of Suburban Life*, Basic Books (New York).

Tabellini, Guido, 2007, 'Culture and Institutions: Economic Development in the regions of Europe',

http://www.igier.uni-bocconi.it/whos.php?vedi=327&tbn=albero&id_doc=177

Temple, Jonathan, and Paul A. Johnson, 1998, 'Social Capability and Economic Growth', *The Quarterly Journal of Economics*, 113 (3): 965-990.

Vico, Giambattista, 1744, *Principi di Scienza Nuova*, Stamperia Muziana (Naples, Italy).

Woolcock, Michael, 1998, 'Social capital and economic development: toward a theoretical synthesis and policy framework', *Theory and Society*, 27: 151-208.

Zak, Paul J., and Stephen Knack, 2001, 'Trust and Growth', *The Economic Journal*, vol. 111, no. 470: 295-321.

Chapter 2

DATABASE

2. I. INTRODUCTION

In chapter 1 the objectives of the project and the theory behind it have been outlined. The project's primary data source is the database constructed by Morris and Adelman (1988). This chapter describes in detail and revises the data, presenting the new version of the database and discussing the amendments made.

The chapter is organised as follows: Section 2. II briefly describes the database and reviews some suggestions for improvement that have been made in the literature. Section 2. III gives a detailed account of the variables in the data, their source, and the precise construction of any derived variables.

2. II. INTRODUCING THE DATABASE

Morris and Adelman (1988) provide an extensive socio-economic database for a wide range of countries for the period 1850 to 1914. This highly valuable database has not received the attention it deserves. In this section, we will first give an overview of the database (subsection 2.II.1); a second subsection (subsection 2.II.2) discusses some specific comments that it provoked.

2. II. 1. OVERVIEW OF THE DATABASE

The basis of the data used for the construction of the Social Development Index for the late nineteenth century is a database provided by Adelman and Morris. The extensive data appendix accompanying their 1988 book summarises an investigation that lasted more than twenty years. It contains cross-sectional data for 23 countries from all continents and referring to 35 summary variables. These variables depict the socio-economic structure of the observed countries between 1850 and 1914. Three sub-periods are distinguished in the original database: 1850-1870, 1870-1890, and 1890-1914. Cross-sectional data are supplied for every sub-period. The variables in levels and proportions refer to the initial year of each period, while those capturing change or characteristics refer to the whole 20 year period.

A complete list of variables is as follows:

Economic variables:

- level of per capita income
- change in per capita income
- percentage of labour in agriculture
- abundance of agricultural resources
- growth of exports
- shift in structure of export sector
- change in average real wages in industry
- change in average real wages in agriculture

Technological variables:

- development of techniques in industry
- improvement in industrial techniques
- improvement in industrial techniques, lagged
- development of techniques in agriculture
- improvement in agricultural techniques
- improvement in agricultural techniques, lagged

Transportation variables:

- development of inland transportation
- improvement in transportation, lagged

Demographic variables:

- total population
- population growth, lagged
- net immigration

Educational variables:

- illiteracy
- spread of primary education, lagged

Land institutional variables:

- predominant form of land tenure
- Concentration of land holdings
- Favourableness of land institutions

Social variables:

- Urbanisation
- Attitudes toward entrepreneurship

Political variables:

- Economic role of government
- Character of political leadership
- National representative institutions
- Political stability
- Foreign economic dependence
- Colonial status

Market institutional variables:

- Development of market institutions
- Spread of market institutions
- Spread of market institutions, lagged

The purpose of the classification above is to facilitate further discussion. The classification does not play a role in the calculations of the main index, although this classification is used later in the thesis for some sub-group calculations. The classification of the authors has been modified: Their "economic" variables are divided in this thesis into three categories: transportation, technology, and purely economic; while Morris and Adelman (1988) do not make this distinction and place all these three under a single category instead. As to scaling, all variables in the Adelman and Morris' database are categorical; every category is represented by a letter ranging from A (top category) to I (bottom category) , a set of potentially nine categories. The number of categories ranges from 4 to 9, depending on the variable. For more subtle distinctions between two categories, occasionally, we find the plus (+) or minus (-) sign added to the label; for instance, A+ or B-. A detailed variable-by-variable description is given in section III below.

The authors used this database to explore the different patterns of economic development amongst sub-groups of countries between the mid-nineteenth century and the First World War. A characteristic feature of this database is that, apart from purely economic measures, it contains a wide range of political, institutional, and social indicators. This feature makes the data particularly suitable for the construction of a Social Development Index (SDI).

Apart from containing a wide range of variables that are of interest here, the database also contains a wide range of countries including all continents. In particular, both

relatively rich and relatively poor countries are included, from different geographical locations. The inclusion of poor countries is crucial to avoid cross-country regression selection bias. Selection bias occurs if subjects (in this case: countries) in the sample used in a statistical analysis are special in some characteristics that are unobserved to the researcher. As a result, the subjects in the sample are not representative for the population under study. An archetypical example of selection bias is the catch-up growth study based on OECD statistics. Inclusion of OECD countries alone introduces a selection bias. The sample of countries in the Adelman and Morris database is diverse enough to believe that there is not a problem of selection bias. It is important to note that not all variables for all countries are observed for all years in the period. As a result, the set of countries for which complete data are available may well be a non-random sample from the original database. Adelman and Morris avoided this potential problem by using a wide enough time intervals of 20 years instead of a one-year time periods. For the second half of the nineteenth century, constructing a measurement for a period of 20 years is feasible for a large set of countries, thus allowing for the inclusion of a good number of subjects. The countries they eventually include vary over a range of development stages and geographical regions, thereby avoiding selection bias.

The list of countries is as follows:

1 Argentina
2 Australia
3 Belgium
4 Brazil
5 Burma
6 Canada
7 China
8 Denmark
9 Egypt
10 France
11 Germany
12 India
13 Italy
14 Japan
15 Netherlands
16 New Zealand
17 Norway
18 Russia
19 Spain
20 Sweden
21 Switzerland
22 United Kingdom
23 United States

2. II. 2. POLLARD'S REVIEW ON THE DATABASE

Sidney Pollard published a review of Morris and Adelman (1988), in which he describes some potential sources of improvement. This section discusses Pollard's criticisms to Morris and Adelman (1988).

Pollard sympathises with the effort to measure what had always been vague theorisation (Pollard, 1988:699), and acknowledges the need to make explicit what development theories say implicitly (*op. cit.*:670). The most valuable contributions of his review are suggestions for improvement of the database. We find three of them particularly useful for further discussion:

Firstly, the appropriateness of using the country as the geographical unit of measure is questioned. Can small countries like Belgium be compared with large countries like Russia or China at the same level? Given that industrialisation was more regional than national, the optimal geographical unit of measure seems to be the region instead of the country, especially for large countries. Taking the country as the unit of study implies that regional particularities are compounded into the national average, neglecting interesting regional features.

The limitations imposed by taking the country as geographical unit of analysis are acknowledged by Morris and Adelman (1988:62) and almost universally present in statistical development studies. Although regional differences have been fruitfully exploited in many economic historical case studies, regional data is simply not generally available for all the variables, countries and periods to be considered in the analysis. Also, using the country as geographical unit of analysis allows comparing the results with related work that generally uses countries as units.

Secondly, Pollard disagrees with assigning scores to political and organisational characteristics, which implies that Adelman and Morris would find some forms of organisation better than others. I find ranking to be fundamental for the Adelman and Morris' investigation, because their aim is precisely to explore whether some socio-political organisations might be more favourable to economic development than others. Likewise, ranking political and organisational characteristic is fundamental for this

investigation as well. Suppressing scores would completely change the nature of the investigation. There is no need to impose an *a priori* hypothesis, as Pollards' critique implies; it is the statistical technique which distinguishes whether those characteristics are positively or negatively associated to other variables in the database. Furthermore, political indicators have been incorporated to the development literature during the 1990's. To sum up, the introduction of qualitative variables into a quantitative study is enriching, not impoverishing.

And thirdly, Pollard finds a bias in the choice of variables towards the institutional ones. The Adelman and Morris database has a large number of institutional variables, compared to other international historical databases like those of Mitchell (2003a, 2003b, 2003c) or Maddison (2003). I agree with Pollard that in a study that aims at exploring patterns of economic development, the inclusion of many institutional variables biases the result towards the importance of their role. In this study, conversely, the ability to develop institutions is crucial to understand the level of development of the society, and, therefore, serves for the purpose of contributing to construct a social development index.

As a conclusion to section II, the introduction to the database, one can say that the Adelman and Morris database has unique characteristics from which an economist looking for social influences in historical perspective can certainly profit. These are: The database describes the situation of the economy in the late nineteenth century in conjunction with a detailed picture of the institutional framework, and some interesting social attitudes in different countries around the world.

2. III. REVISION OF THE DATABASE

This part of the chapter explains in detail the revision of the data, including re-codification of the entire database, criteria used in data selection and other data transformations, and discussion of some of the more broadly used additional sources. The last and longest part is an extensive variable-by-variable description of the data.

2. III. 1. DIGITALISATION, TEMPORAL DIMENSION AND CODIFICATION

In order to carry out a statistical analysis, it is necessary to digitalise the database. Conversations with Irma Adelman revealed that the existing digital versions of the database are now lost. Moreover, the disappeared old files were written in now obsolete formats. Thus, the task of recovering the database in currently operative digital format is essential. The way this task has been done is typing the data into a computer directly from the published paperback. Some considerations have been taken into account and are explained below.

Regarding periodisation, recall that three sub-periods are distinguished in the original database: 1850-1870, 1870-1890, and 1890-1914. Cross-sectional data are supplied for every sub-period. The new database creates time cuts at the 4 limiting years, with the exception of the last period, which end has been replaced by 1910 for two reasons: First, to give homogeneity to the time periods (all of 20 years); and, second, because of convenience in data sourcing other than Adelman and Morris. So, the time cuts are 1850, 1870, 1890, and 1910. When extracting information from this database, it has been taken into account that the variables in levels and proportions refer to the initial year of each period, while those capturing change or characteristics refer to the whole 20 year period. When assigning a value referring to a variable expressing change or a characteristic, it refers to the preceding 20 year period.

In addition to the data corresponding to the 4 limiting years defining the three sub-periods appearing in the final tables of Morris and Adelman (1988), it has been possible to add data for an additional year, 1830, for a limited number of variables. The variables available for 1830 are those for which the database incorporates a lag referring to the preceding 20 year period. These are improvement in industrial techniques, improvement in agricultural techniques, improvement in transportation, population growth, spread of primary education, and spread of market institutions. Although they have been compiled, these are not included in the calculations due to limitations in the number of variables available.

The final format of the data in Morris and Adelman (1988) is a list of 35 variables for every one of the 23 countries and the 3 sub-periods aforementioned. Every one of these variables has an ordinal codification ranging from A to D, A to E, etcetera, until A to I,

depending on the number of categories of the given variable. The number of categories of every variable is arbitrarily assigned in the Adelman and Morris database and depends on the degree of possibilities to differentiate between categories, according variable-specific characteristics. The number of categories of every variable is explained in detail in the list of variables below one-by-one.

It is important to note, though, that, despite the fact that the number of categories has been kept the same and varies from variable to variable, the range of variation is now the same across all categorical variables in the database. This has been achieved thanks to a change in the codification of all variables into a uniform one. Indeed, exploring the codification in more detail reveals that Morris and Adelman (1988) initially used a numerical scoring scheme from 0 to 100 running parallel to the alphabetical scale, and then translated numerical scores into letters in order to make the variables categorical. This facilitated the allocation of countries in one category or another. Sometimes, when a country was falling in between two categories, pluses or minuses were assigned (A- or B+). However, for the purpose of this piece of research there is no need of grouping countries into certain categories prior to the statistical analysis. As long as they have an ordinal score for every variable that is enough. For this reason, it is feasible to recuperate the parallel numerical scale, which runs from 0 to 100 and gives uniformity in the range of variation across all variables as well as increases precision and facilitates the statistical analysis.

2. III. 2. GENERAL CRITERIA USED TO SELECT, MERGE AND TRANSFORM DATA

Incorporating information to the Morris and Adelman (1988) database proceeding from other international databases requires setting some homogeneous criteria. We dealt with the following difficulties: source selection, changing country boundaries, heterogeneity of definitions, and missing data.

The fact that the database is constructed from several sources comes with some associated problems. All sources have their weak and strong points. One simply would like to select the best variables from each of the databases but it is hazardous to combine these data sources. The main criterion for choosing the preferred source is reliability (see discussion of sources below). Once chosen, we stick to a single source for each

variable.

Another important topic is re-labelling of countries and removal or modification of some borders, which makes some countries require special attention and care in data entry. Fortunately, the main sources of data are intended for long-term historical comparisons, so they already supply data on a constant-borders basis (see, for instance, Maddison, 2003). In our database, Russia and the United Kingdom have been considered a single unit throughout; West and East Germany have been unified into Germany throughout. Burma or Myanmar is referred to as Burma throughout.

Another difficulty in data merging is heterogeneity of reporting criteria. The same information may be reported in different ways. For instance, the gross domestic product (GDP) of a particular country may be presented in *per capita* terms in one source and in levels in another, or it may be expressed in current prices in one source and indexed by a constant base-year in another. Another example is the varying definition of external trade, which depends on whether goods in transit where reported as trade or not, amongst other things. In the literature, Mitchell and Maddison are particularly aware of these complications and are careful to state their choices with clarity. We have paid special attention to the homogeneity of reporting criteria. Choices are clearly stated in each variable's description.

Finally, some variables are not available for certain countries and/or years; therefore a blank is created. Our policy has been making the most out of the data, within reasonable limits. In particular, when a data was missing for a given year it has been replaced by the one corresponding to the closest year available, unless this year happened to be closer to he next reporting year. When an intermediate year is missing in a series that is complete or quasi-complete, or which does not show drastic changes, we calculate an estimate by interpolation of the preceding and following figures, assuming a constant rate of growth. Finally, for the cases where no close alternative year was available or the series did not show enough continuity and stability for interpolation no account has been reported.

2. III. 3. ADDITIONAL DATA SOURCES

Information from several additional sources has been incorporated into the database. Since the database was published in 1988, a substantial amount of new material has appeared. It is worthwhile to describe the main available international historical statistics that have been explored for updating this database:

Maddison

Angus Maddison, in his book *Monitoring the world economy, 1820-1992,* provides a broad array of estimates which describe 'virtually the entire world over the period from 1820 to 1992' (Maddison, 1995:15). The dataset includes basic macroeconomic variables: population, GDP, trade, government spending, education, transportation, employment, working hours, and productivity. An advantage is that 'the estimates are adjusted to eliminate the influence of changes in national boundaries', which is of special interest for long-term studies (Maddison, 1995:17). The president of the OECD Development Centre in 1995, Jean Bonvin, tells us in the preface that these data are of particular interest for economic historians and growth economists, since they document long-term growth trends. This makes this database especially suitable for this project.

Maddison provides some information about the reliability of the data. He reveals that 'amongst OECD countries, the weakest long-term GDP estimates are for Greece, Ireland, New Zealand, Portugal, Switzerland and Turkey' (Maddison, 1995:17-18). Of these countries, only New Zealand and Switzerland are in our sample. For China, Maddison offers a separate case study in *Chinese Economic Performance in the Long Run.* His more recent book, *The World Economy: A Millennial Perspective*, initially intended as a revision of his earlier work, eventually exceeded the status of revision; it includes many more series and goes back much farther in time. We used Maddison's data for the series on per capita GDP and population statistics. We also considered the suggestions for improvement of GDP measures given in Leandro Prados (2000), which entail a different Purchasing Power Parity (PPP) approach. Eventually we use the estimates from Maddison. More on this in the separate section on GDP below.

Finally, *The World Economy* (Maddison, 2006) brings together the two volumes on the world economy: Volume 1: *A Millennial Perspective* and volume 2: *Historical Statistics*

(Maddison, 2001 and 2003 respectively).

Bairoch

The book edited by Bairoch and Levy-Leboyer (1981), *Disparities in Economic Development since the Industrial Revolution* provides a limited number of tables of comparative statistics illustrating international inequalities in development. The book devotes special attention to Third World countries and former colonies for varied periods of time, which is a statistical rarity. This is also true for other publications by Bairoch, like *Diagnostic de l'Évolution Économique du Tiers-Monde, 1900-1968* (English version entitled *The Economic Development of the Third World since 1900*).

Bairoch's data are not completely reliable, as the process of calculation is not completely spelled out. In this respect, Maddison's data can be considered superior. Nevertheless, Bairoch computes industrialisation levels for the 1970's and 1980's that are not currently available from any other source. Similarly, *La Population Active et Sa Structure* (Bairoch, 1968) continues to be the main source for the labour structure of the active population by sectors, as it is shown in Mitchell (2003a, 2003b, 2003c; see for instance Mitchell, 2003a:143).

Mitchell

A major source for our economic variables has been the regularly updated International Historical Statistics compiled by Brian R. Mitchell. Mitchell gathers a wide range of variables referring to population, migration, wages, sectoral production figures, and transport and communications, among others, dating back to 1750. The last edition of the three volumes, covering respectively Africa, Asia and Oceania, the Americas, and Europe, were published in 2003 (Mitchell, 2003a, 2003b, and 2003c).

We have used Mitchell's statistics for updating most of the economic variables in the database. There are three main advantages in using Mitchell as a data source. The first advantage is that his statistics are regularly updated, as mentioned above. The second advantage is reliability, as official sources were used. Finally, he approaches the data collection with the necessary care and precaution. To illustrate the latter, let us cite some examples:

> 'Statistics are used nowadays (...) as a major raw material of much economic history, especially of economic growth. (...) [T]he subject is by its very nature concerned with quantities, (...) though no sensible historian would ever claim that statistics can tell the whole story.' (Mitchell, 1998a: vii).

Mitchell acknowledges that his data were conceived to be an answer to the increasing interest in comparative development, and provides long time-series into a comparable format. He is very careful not to fall in common pitfalls, like problems of definition. His awareness of the problem of 'the existence of data which *seem* to relate to the same things in different countries or at different times, but which do not in fact do so' is illustrated in the next examples (*ibid.*):

> '[I]n some times and places exports include bullion, in others they do not. Pig iron can include or exclude ferro-alloys; bank deposits may include those of other banks or they may not; corn output can be measured by volume or by weight; and so on' (*ibid.*).

Thus, Mitchell shows awareness of heterogeneity of definition problems across space and time, while promising to warn us when these occur. He is equally aware of possible understatement or overstatement of some statistics, due to some perverse incentives not to reveal the information truthfully, as in the cases of 'registration of one's true age if one was a young man liable to military service, and the smuggling of dutiable imports'(*op. cit.*:viii). Mitchell claims that, in general, statistics become more reliable at the end of the nineteenth century, when some countries began collecting statistics for its own sake, not with any direct fiscal or military aim. The first century of Mitchell's data, corresponding to 1750-1850 should hence be considered less reliable, but falls outside our period of study. For our period of study (1850-2000), in general, we consider Mitchell's data reliable and up to date, and we have relied heavily on them for the reconstruction of many economic indicators.

3. III. 4. VARIABLES

A discussion of all variables in the database follows, detailing how they were constructed as well as describing the changes we incorporated, if any. In order to facilitate comparison with Morris and Adelman (1988), the classification of the

variables used in this section corresponds to the one used in their data appendix (see in Morris and Adelman, 1988:223-5).

Economic influences

Many of the variables are under the title "economic influences", thought they do not refer to purely economic aspects, but also to technology adoption and improvements in transportation. A list of all the economic-related variables in the database follows:

- Level of per capita income
- Rate of change in per capita income
- Percentage of labour in agriculture
- Rate of growth of exports
- Degree of shift in structure of export sector
- Change in average real wages in industry
- Change in average real wages in agricultural poor
- Development of techniques in industry
- Rate of improvement in industrial techniques
- Rate of improvement in industrial techniques, lagged
- Development of techniques in agriculture
- Improvement in agricultural techniques
- Improvement in agricultural techniques, lagged
- Relative abundance of agricultural resources
- Development of inland transportation
- Rate of improvement in inland transportation, lagged

We will discuss them in subgroups.

INCOME VARIABLES

Within the economic variables, we have a subset referring exclusively to income. These are:

- Level of per capita income
- Rate of change in per capita income

Per Capita GDP (Level of Per Capita Income)

Gross Domestic Product (GDP) is the variable that we will discuss most extensively, since it has received a vast amount of attention in the literature ever since Adelman and Morris constructed their database circa the 1970s.

When Adelman and Morris built their database during the 1960s, 70s, and 80s, per capita income point estimates would still not be well established for many countries. Income variables were 'substantially changed' over the period between the publication of their first papers until the publication of their book in 1988, due to new investigations in per capita income being published (Morris and Adelman, 1988:225). A lack of consensus on how to deal with the problematic issues arising from the calculation of income, led them to use per capita income brackets instead of point estimates. They made sure that their final income variable was less sensitive to variations simply by taking broad enough income brackets (six in total). Their main sources were Bairoch (1976), Maddison (1979), and Mulhall's *Dictionary of Statistics* (1899). For the countries absent in these publications (Argentina, Brazil, Burma, Canada, China, Egypt, India, and New Zealand) they relied on comparisons of country studies.

Nowadays, point estimates are available for all countries in the sample and much has been published on how to calculate and re-calculate them. Though the debate is still alive, point estimates are no longer the exception but the norm in both cross-country and inter-temporal comparisons. Therefore, although the Adelman and Morris' classification is valid and informative, there is an argument for replacing it by more precise per capita GDP point estimates. Various alternatives are available; each of them having associated advantages and caveats. Maddison's is possibly the most well- established series on international historical per capita GDP estimates. He considers all sources and measurement alternatives available, and discusses them carefully. The variable that we use is based on Maddison (2003). In order to further defend this choice, an exploration of the alternatives follows.

GDP is the most broadly used amongst the income/produce indicators. Most economic databases contain GDP, and it has been progressively adopted as the reference indicator of economic activity amongst economists and economic historians. Surprisingly, even its closest relative, Gross National Product (GNP), is way far less used nowadays. GDP

accounts for the output produced within a given country, whereas GNP refers to produce by *nationals*, not necessarily within the borders. There is no reason to break the tacit consensus, and so GDP is the income indicator we use here as well.

GDP, however, may not always be available, or be only available with considerable measurement error. In this case, Jeffrey Williamson proposes to account for economic activity from the income side. He suggests that examining wages and relative factor prices may give a better idea of economic activity than GDP series in that case. Williamson's data appendices provide series for different regions of the world: The Mediterranean, Asia and Latin America (Williamson, 1999, 2000a, and 2000b).

Given that Maddison provides us with GDP series for a wide range of countries and years, the only reason for replacing it by Williamson's income based approximations is that GDP in Maddison's would be severely misestimated. Actually, even if there *would* be some degree of measurement error, there is no *a priori* need to replace it by another variable that is *also* an approximation. As Williamson points out, wage series can serve as a replacement for GDP when the latter is not available or strongly unreliable, no matter how respectable his series may be.

Moreover, there is an inconvenience in using Williamson's series for the database we are re-constructing here. The use of wages in the calculation of income (and, therefore, the rate of change of income) may complicate the joint statistical analysis with other variables already in the database, namely the rate of change of wages in agriculture and in industry respectively. For instance, any regression analysis that explains a certain variable with all the above mentioned wage measures included as conditioning variables, would suffer from large standard errors of the estimates of the parameters due to multicollinearity[1]. Also, inclusion of all the above mentioned wage variables in a principal components analysis would fail to generate substantially separated income-based dimension due to the strong correlation of this variable with the other two wage variables. This is another reason to prefer here the production approach to GDP.

[1] Multicolineary problems appear when we include into a regression two or more variables that are highly or perfectly correlated with each other. They are, in essence, the same variable; hence their effect on the dependent variable cannot be separately identified.

Now that we have discussed our preference for GDP and, in particular, its production approach, it is necessary to deal with a series of problems that emerge when comparing GDP measures across different countries and time periods, in particular those related to purchasing power parity (PPP) and exchange rate adjustment.

First, we need to make the distinction between nominal and real GDP. Nominal GDP accounts for the total value added in a country during a given year, evaluated at current prices. Since prices change from one year to another, as a measure of production, nominal GDP in one year cannot be compared to nominal GDP in another year without first transforming the two of them into a common price set. Real GDP has been devised in order to allow for comparisons of output over time, by means of evaluating production at a constant price level corresponding to an arbitrary year. Note that the values taken by real GDP depend on the choice of the base year. It is necessary to use real GDP, as opposed to nominal, in order to enable comparisons across time.

The invaluable advantage of having series that are comparable over time comes at a cost. Relative prices of goods and services within an economy vary over time. By using very distant PPPs, in particular those of 1990 as in Maddison (2003), for conversions to comparable real GDP in the nineteenth century, we are implicitly assuming that relative prices of goods and services in say 1890 are the same as those in 1990. Leandro Prados comments: 'Unfortunately, by accepting a distant PPP as the point of reference, the procedure, as pioneered by Bairoch (1976) and Maddison (1982), introduces distortions and ambiguities in inter-temporal comparisons', since the composition of production, consumption and relative prices all vary as growth occurs (Prados de la Escosura, 2000:3-4). The problem is that even if we evaluate production at a constant price set in order to facilitate inter-temporal comparisons, the resulting real GDP might not give us straightforward indication of the standard of living at a given point in time. This is due to the fact that relative prices change over time. For instance, owning a horse in the nineteenth century does not mean the same as owning a horse nowadays, since a horse in the nineteenth century was much more valuable in relative terms than it is today. In the same way, owning a very basic radio set does not have the same relative price in comparison to other goods and services as it had several decades ago. The whole set of ratios of prices of one good to another has changed. Thus, evaluating output at a fixed set of prices can lead to misleading interpretations of the standard of living.

This is a complicated topic to handle, where correcting one bias normally implies introducing another. Biases are an especially delicate issue in comparative studies, since they can lead to changes in the relative position of income of the different countries.

Chained price indices avoid the problem of setting a constant price level corresponding to an arbitrary base year. A new methodology was devised by Summers and Heston (1988) in order to avoid changing both levels and growth rates of all past and present real GDP every time the base year was updated. The maxima for the chained system would be: Write history just once! The rate of change of real GDP between two consecutive years is calculated by taking as common price level the average price level of the two years. Then, the resulting rate of growth is reflecting the *real* change in production. A time series for real GDP is constructed by "chaining" the resulting real rates of growth. This methodology was adopted by the United States governmental economic reports in 1995, and it has been used ever since. It is more complicated to compute than the indices used before 1995, but the GDP series are more stable.

To deal with the associated complexities of the calculation of the chain price index, the Bureau of Economic Analysis (BEA) introduced a 'dollar-denominated real output series that are based on (...) the new indexes but that have the computational simplicity of constant-dollar series' (Landefeld and Parker, 1997). Indeed, '[t]hese annual changes are "chained" (multiplied) together to form a time series that allows for the effects of changes in relative prices and in the composition of output over time' (*ibid.*). The rates of real growth are correct, but we need to establish a base year in any case, from which to start multiplying growth back and forth. For this matter, the BEA introduced "chained 1992 dollars". '[T]hey work well for periods close to the 1992 base year, but they may produce increasingly misleading results as one moves away from that year' (*ibid.*). Thus, in the end, the need for a reference year when calculating GDP levels from the derived growth rates implies that biases cannot be completely avoided, and are worse as we move further back in time.

The source we use, Maddison (2003), incorporates the chained price index for recent United States data, as provided by the official accounts, and brings this methodology back to 1950. He does not bring it further back in time because doing so would imply such big drift from traditional history–especially for the interwar period– that he prefers to be conservative for the moment and maintain the fixed price index for years previous

to 1950. Moreover, no other country but the United States has provided chained GDP for long periods in the past (Maddison, 2003:80).

Rate of Change in Per Capita Income

In the same way six income brackets referred to the level of per capita income (see *Per Capita Income*), Adelman and Morris derived five different categories of change in per capita income, sorted by average annual rate of growth during a given 20 (or 20+) years period. i. e. estimated average rate of growth 1850-70, estimated average rate of growth 1870-1890, and idem for 1890-1914. These categories range from 'probably over 2 percent average annual' to 'decline in per capita income' (see table A3 in Morris and Adelman, 1988:252-3).

Growth estimates are based on fragmentary data. They use several different bibliographical sources for every country, and the placement of countries in the corresponding growth categories are based on a combination and contrasting of different primary (and some secondary) sources such as: GDP, GNP, GNP growth, value added by sectors and by productive enterprises, agricultural yields, movements in gross sectorial output, quantitative exports, growth of exports, indices of purchasing power of exports, the percentage of a particular good (cotton for Egypt) in the total sectorial production and in the total of exports, volume of foreign trade, output of the railway system, tax yields, population, employment, wages, market prices, labour inputs, energy inputs, growth of population involved in manufacturing, and qualitative information. They prove to be well-informed, and they have a good overview picture. Of course, to some extent, they had to make –and accept from sources– some assumptions about the basic structure of the economy, like, for instance, relative productivity levels across sectors or the rate of growth of population. Assumptions tend to be supported by qualitative information.

It stands clear from the previous discussion in *Per Capita Income* that a great deal of preciseness can be gained from using the new income variable discussed above for the calculation of average rates of growth. Again, we move from broad categorisation to point estimates.

THE INDUSTRIALISATION VARIABLES

Within the economic variables, we have a subset of variables referring exclusively to industrialisation. These are:

- Development of techniques in industry
- Rate of improvement in industrial techniques
- Rate of improvement in industrial techniques, lagged

Level of Development of Techniques in Industry

Adelman and Morris account for the 'replacement of preindustrial methods by mechanized methods' with this variable (Morris and Adelman, 1988:97). Their sources are descriptive materials, such as:

- employment in factories
- average size of factories by employees
- output per worker
- value of output by factory
- horsepower per worker
- horsepower per enterprise
- average horsepower in manufacturing
- size distribution of horsepower by enterprise
- relative importance of manufacturing
- concentration of factories
- distribution of enterprises by size

In order to facilitate cross-country comparisons they concentrated on consumer goods and machinery production only. Countries at a given year were allocated to one of seven ranking categories, starting with countries with mechanised textile spinning and weaving, consumer goods' factory production, and machinery industry with interchangeable parts for the most technologically advanced. At the bottom end of the classification we find countries with no factories with inanimate power. Therefore, the classification of countries is based on the modes of production.

Rate of Improvement in Techniques in Industry

How fast is technology assimilated? The speed of adoption of new technologies may depend on several things. It is a matter of capabilities, adequacy, and starting level.

For the Social Development Index, we are interested in the capability element. There is an element of capability, in the sense that a society may be more ready than another to welcome new productive ideas and applications. Recall here the early patent laws of the United States, ahead of its time.

As far as adequacy is concerned, how adequate a certain production technique is to the environment and relative input ratios has nothing to do with capabilities. There are examples of slow technology adoption due to non-adequacy to the environment. For instance, the adoption of the plough in certain parts of Africa was retarded by the non-suitability to the soil, plot features, and required animals.

Finally, other nations rather than England could also have started a productive revolution (not necessarily based on textile industry and the use of the steam engine), with machines and/or production modes adapted to its own environment. The argument is that many underdeveloped countries were rich in natural resources.

So, we shall assume that slow technology adoption is rooted at least partially in capabilities of the society. This is not to say that the geographical aspect is completely neutral. Another variable in the database, –namely the abundance of land (or land/labour ratio)–, is capturing (though imperfectly) the initial endowment of natural resources.

A much less neutral aspect is the starting point of technology. A country is likely to incorporate new techniques faster, the least is the initial level of technology, since the difference between the actual level and the potential level of the leader creates a gap for catching-up. Indeed, industrial techniques were incorporated *faster* in Germany or the United States than in the United Kingdom, especially at the end of the nineteenth century, thanks to the existence of a catching-up gap. So, the United Kingdom should be given more credit for adoption of new techniques in industry for breaking ground. This is done by taking into account the *initial* level of development of techniques in industry

at the beginning of every 20 year period (see previous variable), which gives more relevance to countries that worked with the new production techniques first, i. e. the United Kingdom. Adelman and Morris already took into account the catching-up effect in the technological growth variable by ranking lower countries that started from a narrower base.

Their rate of improvement of techniques in industry 'measures the rate of expansion of mechanised industry using inanimate power', as explained by Morris and Adelman (1988:98). Countries where classified in seven categories, the best one being substantially industrialised countries (above 25 percent of industrial output) with an annual rate of growth of industrial output above 3 percent, of which most was factory production. Belgium, Germany, the United Kingdom and the United States qualified for this category in one or more periods. Countries at the bottom of the classification only experienced hand-manufactures or artisan industry growth, if any at all. Argentina, Brazil, Burma, India, and New Zealand around mid-nineteenth century fall into this category. Information on which to base their classification was based in numerous sources and varies for every country.

Rate of Improvement in Techniques in Industry, lagged

Adelman and Morris also considered a lagged version of the same variable for their comparative development purposes. This has not been incorporated into our analysis because of a lack of reasons to include it, especially given that the contemporaneous levels are already included in the database. Nevertheless, the information in this variable has been helpful to the extent that we have used the data corresponding to the period 1830-1850 to further document the rate of improvement of techniques in industry.

THE AGRICULTURAL VARIABLES

This set of variables encompasses the level and rate of improvement of agricultural techniques. However, the focus is not only techniques, as in the case of industry. The agricultural variables are:

- Level of development of techniques in agriculture
- Rate of improvement in techniques in agriculture
- Percentage of the labour force in agriculture

- Relative abundance of agricultural resources

Level of Development of Techniques in Agriculture

The "level of development of techniques in agriculture" is a four-dimensional variable which captures the extent to which modern agricultural techniques were (or not) used. These techniques are:

- Horse-drawn iron or steel machines
- Improved crop rotations
- Irrigation
- Fertilisers

All these techniques increase productivity, but some more than others. Adelman and Morris assigned different weight to each of them according to the probable contribution to productivity. They explain the ranking of importance in great detail. Some fragments follow:

> 'Greatest importance was given to the adoption of labor-saving machinery. We concentrated on machinery used in harvesting since harvesting was usually the major production bottleneck. Improved iron and steel plows were weighted next most heavily (...). Next in importance were crop rotations, seed selection, irrigation systems, and fertilizers [apparently, all equally valued] (...). Weighted third most heavily was the spread of crop diversification and mixed animal-crop farming (...). The final type of improvement, the use of animal farming of enclosure, stock breeding, and supplementary feeding, was given significant weight in ranking countries where pastoral farming was very important' (Morris and Adelman, 1988:226-7).

This detailed ranking of importance, paired with the information that existing literature revealed about whether a certain technique was or not utilised, give rise once more to a seven-folded categorisation. There is not evidence of them assigning specific figures as weights. It seems that they just had in mind the described ranking of importance when assigning a given country-year observation to a certain category. Therefore, it is difficult to contradict them, and also to reproduce their carefully studied decisions. They relied strongly, if not exclusively, in specific country literature. Once they finished their categorisation, they double-checked their results the previous comparative studies by

Mulhall (1899), and Clark (1957). Thus, the Morris and Adelman (1988) classification for the level of development of techniques in agriculture is very careful and quite reliable.

Rate of Improvement in Techniques in Agriculture

The rate of improvement of agricultural techniques is a variable with six categories, ranging from fast adoption of labour-saving machinery that significantly raised productivity to negligible improvements in a 20 years period. This variable is based on the rate of change of the variable above (level of development of techniques in agriculture).

Percentage of the Labour Force in Agriculture

The original variable for the percentage of labour force in agriculture in Morris and Adelman (1988) has various problems. On the first place, they were dealing with estimates or guesstimates in most cases; 'few are based on properly constructed census data' (Morris and Adelman, 1988:227). Secondly, there was not a homogeneous definition of active population across countries. So, the authors present a table of estimates of labour force in agriculture which they did *not* use in the analysis, since it was heterogeneous and incomplete. Again, Adelman and Morris show common sense in not wanting to force point estimates when the quality of the data did not allow for it. Instead, as in per capita income, they used percentiles. i. e. percentage of agricultural labour force' brackets. These brackets are:

More than 70 percent of the labour force engaged in agriculture,
61-70 percent,
51-60 percent,
41-50 percent,
31-40 percent,
21-30 percent, and
20 percent or less of the labour force engaged in agriculture,

giving rise to one category each. They used various sources; Clark (1957), Bairoch (1968), and Cipolla (1974) are widely referred to.

Since Adelman and Morris constructed their database, there is enough new material to allow for a full update of this variable. The most recent series of international historical statistics by Mitchell (2003a, 2003b, and 2003c) gather point estimates of the percentage of economically active population in the primary sector of the economy (agriculture, forestry, and fishing) starting in the mid-nineteenth century until nowadays for many countries all over the world, including the countries in our sample. These stem as far as possible from official censuses and are harmonised to a greater extent than those available at the time Adelman and Morris constructed their database. It is worth noting that Mitchell's statistics until the 1960s are mainly based on Bairoch (also consulted by Adelman and Morris), and he puts through the same complaints as Adelman and Morris about the series: multiple nomenclatures and lack of clear separability between agricultural and industrial working class for early periods. However, Mitchell emphasises that Bairoch and his team 'made every possible effort to achieve international and intertemporal comparability' (Mitchell, 2003c:143). Moreover, Mitchell says, 'the degree of precision has improved over the last forty years' (*ibid.*). The biggest advantage of using Mitchell's statistics instead is that they should bring, in principle, much more preciseness to the variable. Categories have been removed and replaced by point estimates.

Methodologically, Mitchell does not provide percentage data but totals in thousands of people. The percentage of active population working in the primary sector has been calculated by adding up all sectors and then calculating the proportion corresponding to the primary sector. In some cases, male and female data are available separately; these have been aggregated. For those years where Mitchell provides no data, the closest data point has been assigned instead, unless it is closer to the next period, in which case this source is not used. For those cases where Mitchell's data was not available, the old Adelman and Morris point estimate has been used, including those for 1914, which they made available but were not included in the final tables for further study.

Analysing the result, few additions have been made, mainly modifications of existing data. Furthermore, for most of the countries, statistics remain practically the same. The resemblance between the data provided in the original tables in Morris and Adelman (1988) and Mitchell (2003a, 2003b, 2003c) is remarkable. So, in this case, updating the information available for this variable was more a ratification than a radical change in numbers. This fact adds reliability to the original data.

The resemblance in figures is due to the fact that both Mitchell, on the one hand, and Adelman and Morris, take the data from the same source: Bairoch (see for instance Mitchell 2003a:143 for reference to Bairoch, 1968). In some cases the new data reflect a more advance stage of de-agriculturisation than the old data do. However, these small agricultural labour percentages go down and again up in the new data calculated from Mitchell, which is highly implausible. This may be due to the fact that some of the labour force percentages for some of the sectors are absent in Mitchell's tables due to the lack of information rather than due to the narrow working force. For this reason, the old Adelman and Morris agricultural labour proportions seem more plausible. Thus, the point estimates taken from Adelman and Morris are preferred. Their unused tables showing point estimates substitute the bracket-estimates in the new variable.

Relative Abundance of Agricultural Resources

Many africanists refer to the land/labour ratio in their discussions about economic development. The same is true for economic historians referring to the success story of the United States and other resource-rich economies. In both cases, they point out how relevant the relative abundance of land is. Nevertheless, it is worth noting that in the first case, it is used to explain failure while in the second case is to explain success. Which one is correct, then? In some cases, abundance of agricultural resources has turned countries into resource-exploited colonies, whereas in others it brought prosperity to their inhabitants. Which one occurred was in many cases a matter of politics and institutions, which we will deal with other variables. The relevant point here is that the abundance of resources matters.

Adelman and Morris produced five categories for the abundance of agricultural resources. With this variable, they intended to capture the relative abundance of farmland with respect to the size of the population. We did *not* use their categorisation, but rather the exact figures of the ratio of population to farmland that they provide, measured in population per square kilometre of standard farmland. The data are adjusted for quality of land. Units of "standard farm land" were defined by Colin Clark (1957) as units of land in the principal temperate or subtropical climate farming regions of the world. Then, land in other regions is weighted according to climate characteristics, in accordance with the potential of agricultural productivity (see Clark, 1957:305ff).

Data on farmland correspond to the 1930s, being the climatic classification of the Earth that of Thornthwaite (1933). Farmland is assumed to be constant throughout the period, while population is adjusted.

THE TRANSPORTATION VARIABLES

Two variables in the database:

- Level of development of inland transportation
- Rate of improvement in inland transportation, lagged

Level of Development of Inland Transportation

This variable captures long and short distance established transportation networks. It represents the ease of mobility of people, which is essential to facilitate communication, connections, links of all sorts. The main sources are country studies. The sources used most across countries are Clapham (1936), Clough and Cole (1946), and Girard (1965). Adelman and Morris contrasted their classification of countries based on descriptive information with statistics in order to make sure they coincided. They emphasise that they carried substantial additional research in almost all countries in order to revise the classifications.

All the information compiled was then utilised in order to classify countries according to three criteria, by order of importance:

- long-distance transportation between large urban centers
- interregional networks
- all-weather feeder transportation within agricultural regions

Provided transportation networks existed within each one of these criteria, they also examined the means of transportation, giving more importance to railways than to waterways, and more importance to waterways than to roads. Within roads, they gave more relevance to those suitable for carts or carriages than those only suitable for pack animals. Five categories arise from this classification. They can be succinctly summarised as follows, from best to worst:

- Countries where the major urban centers and ports were linked by railways, all-weather roads and improved waterways. The agricultural sector was well served.
- Railways were less important than in the first category.

- Some populated area of the country was not served by the transportation system, roads for agricultural goods less well maintained, and institutional maintenance was not secured.
- In addition to the previous category, railways did not provide links for internal trade.
- In the last category, most of the populated areas were not well served, and the main transportation system were pack animals.

The alternative to Adelman and Morris is to account for specific amounts of laid railway lines, or transported passengers, instead of assigning degrees of ease of mobility. Instead of assigning every country and year to one of five categories, as Morris and Adelman (1988) did, there is also scope for assigning a precise numerical value linked to specific transportation dimensions. Mitchell (2003a, 2003b, and 2003c) presents a collection of seven transportation related indicators such as length of railway lines, or number of ships registered, which offer the possibility of assigning a figure instead of a belonging to a category. The transportation variables available from Mitchell are:

- Length of railway line open
- Freight traffic on railways
- Passenger traffic on railways
- Merchant ships registered
- Inland navigation traffic (only available for Europe and North America).
- Motor vehicles in use
- Commercial aviation for Europe, and civil aviation traffic for the rest of the World.

The reporting of most of these variables starts before 1850; (we needed some statistics back from 1830 for the calculation of the variable below 'rate of improvement in use of transportation' –see 'rate of improvement in inland transportation, lagged'). We exclude inland navigation traffic because data are only available for Europe and North America. So, water transportation is represented by the registered number of ships only. All the indicators would need to be in comparable units of measure so that they could be put side-by-side. The units of measure are kilometres for the length of railways, both millions of passengers and millions of passenger kilometres by rail, total number of merchant ships registered (all: sail, steam, and motor), thousands of motor vehicles in use (both private vehicles and commercial vehicles), and millions of passenger kilometres in aviation. It is a complicated task to confer homogeneity to a single

indicator derived from all these different numerical scales. Of course, the number of motor vehicles, and specially aviation traffic start only well into the twentieth century. The indicator of motor vehicles in use is incorporated *de facto* in the 1914 cross-section, while the aviation indicator can only be incorporated from the 1929 cross-section onwards, which lays outside the period under study.

In this case, the categories devised by Adelman and Morris provide a more complete evaluation of the inland transportation possibilities. So there is no need to restrict ourselves to one or several punctual measures referring to a specific transportation system.

Rate of Improvement in Inland Transportation, lagged

The original variable for the rate of improvement in inland transportation is a six-categories ranking classification for "breakthroughs" in the level of development of transportation systems. A breakthrough occurs when there is a drastic drop in the cost of travel. This happened, for instance, with the arrival of the railway, or the emergence of mass transatlantic shipments. Sources are different for every country, but the most important ones are Clough and Cole (1946), Mitchell (1973), and Girard (1965); these sources were revised and double-checked by Morris and Adelman (1988).

THE EXPORT VARIABLES

Morris and Adelman (1988) provide us with two variables referring to international trade; i.e., how much movement of goods and services there is with other countries. We are interested in two aspects:

- Rate of growth of total real exports
- Shift in structure of export sector

Rate of Growth of Total Real Exports

The rate of growth of real exports in the original database captures the degree of incorporation of national products into an internationalised economy. Adelman and Morris provide several tables summarising their findings on the probable average rates of growth of real exports and exports in current prices for the 23 countries in the sample

during varying intervals of time. These tables are worth appraisal, especially given the difficulties and uncertainties in calculation.

Difficulties in calculation arose principally from the uncertainty of price levels. Heterogeneity in the quality of price data affects the quality of the variable real exports via the adjustment of current values. In addition, there were some differences in the definition of exports. Definitional differences came from whether goods only partially produced in the country were included or not, and whether goods only in transit to another final destination were included or not[2]. As a result, the estimates would suffer from inconsistencies. For these reason, only four categories would be established. The classification is as follows, by order of probable average annual rate of growth of real exports during a given 20 year period:

over 4%, departing from a moderate or large base

over 4%, departing from a very small base

between 2% and 4%

less than 2 %

As we can read from the categorisation, they took into account the departure point of a country, giving more relevance to growth once the country would already be established as an international exporter.

On top of the problems of calculation described, we encounter a major conceptual disadvantage in using the original Adelman and Morris' rate of growth of exports. This is the fact that their indicator refers just to one side of external trade, i.e. exports; but does not account at all for the other side of external trade, namely *imports*. The way to solve these problems is taking an aggregate external trade indicator which accounts for both imports and exports, and is the least dependent possible on price levels.

Mitchell provides data on the current value of the aggregate external trade for every country. He also focuses on "special" trade only; this is, excluding goods that are in the country only in transit to another country. Differences in the definition of "special"

[2] On the one hand, "special" trade refers only to goods that are either for internal use –in the case of imports–, or produced, at least to some extent, within the country –in the case of exports–. On the other hand, "general" trade includes all the goods entering and leaving the country. Adelman and Morris wanted to focus on "special" exports, but this was not always possible due to variations in the definition of exports.

trade remain, but Mitchell does not seem as concerned about it as Adelman and Morris did. He argues that these sources of variation are unlikely to be of 'great significance' (Mitchell, 2003c:569). Instead, he shows a greater concern with the inability to capture smuggling, especially until mid-nineteenth century, where there was a generalised movement towards free trade. Therefore, the period of greatest concern is not affecting our sample. In Mitchell's words: 'By the time tariffs were once again raised, policing was much more effective than it had been earlier' (*ibid.*). Mitchell considers his series reasonably accurate from the middle of the nineteenth century onwards. His figures are expressed in millions of local currencies; for instance, the current value of aggregate external trade of the Netherlands in expressed in millions of gulden, etc.

One advantage of using Mitchell's series is that both sides of international trade, imports and exports, are reported. So, there is the possibility of including both imports and exports in a joint measure for international movement of goods and services. Potentially, this could improve the reliability of the variable, enriching the meaningfulness (up to which extent a country at a particular moment in time is integrated into the global economy is reflected by both imports and exports). However, Mitchell's data are expressed in current prices, and do require to be weighted by current GDP in order to achieve a comparable measure. This is, in principle, the idea of such an index; but current GDP series are far less complete than desirable. So, unfortunately, using Mitchell's series on current GDP would create too many blanks to be operative. On the other hand, Maddison provides historical series for the value of exports at constant and comparable prices (million 1990 dollars) from 1820 (Maddison, 2006:360). The drawback is that only selected years are available. This can be solved, however, by interpolating. A constant annual rate of growth has been applied. Thus, Maddison's constant prices per capita GDP, as appears in the variable INCOME, has been used instead, together with Maddison's series for the value of aggregate exports.

If we divide the aggregate exports by GDP, we will get a measure of **exports relative to GDP** which is as least dependent on the level of prices as possible. The main advantage of using the new indicator is that dividing the value of exports by GDP makes it independent of price levels or changes in prices. Constant GDP series have been extracted from the same source for consistency. It is necessary to say though, that the new trade indicator has still *some* degree of dependence on price variations. Some inaccuracy in this measure could arise from the price structure of internationally traded

goods with respect to non-traded ones. This potential source of inaccuracy has not been corrected. In other words, we assumed that the deflator of traded goods and that of locally produced goods are the same.

Shift in Structure of Export Sector

During the second half of the nineteenth century (and thereafter), there was a change in the structure of demand from primary to manufactured goods. The "shift in the structure of export sector" measures the responsiveness of national producers to this change in the structure of demand. With the information available from specific country sources, Adelman and Morris classified countries at a given period under one of these four categories:

- Very strong shift in the structure of the export sector (above 10 percent change)
- Moderate shift (between 5 and 10 percent change)
- Slight shift (between 2 and 5 percent change)
- Negligible or no shift (less than 2 percent change)

These are mainly but not exclusively based on statistical data about the structure of exports at the beginning of each period, i.e. 1850, 1870, and 1890. When data were not satisfactory, descriptive sources were used. The breaking points between categories seem to be arbitrary.

The shift in the structure of the export sector accounts for how fast every country was able to adopt the patterns of modernisation in the exports' structure.

THE WAGE VARIABLES

For the application concerning this project, namely finding differentials in social capabilities, real wage levels are satisfactory enough to provide us with a comparison of living standards across countries and time, beyond GDP. On the one hand, a completely new variable has been incorporated into the database, making use of recent literature, namely a real wage index. On the other hand, the wage variables in the original Adelman and Morris database have been kept since they do not make reference to wage levels; they focus on the movement of wages instead. Apparently, whether real wages were going upwards or downwards gave them more an indication of economic development than the level itself. Nevertheless, they provide no justification for this

decision in their 1988 book, although it might be due to the difficulty to harmonise real wage levels across borders. Interestingly enough, they made the distinction between industrial and agricultural wages. The two wage variables in the original database are the following:

- Direction of change in average real wages in industry
- Direction of change in average real wages or income of employed agricultural poor

These two variables give an indication of how fast was purchasing power climbing. So, we may want to preserve them in the data analysis, because they give an idea of the differentiated evolution of wages in industry and agriculture, which is still difficult to substitute by nowadays historical statistics.

Real Wages

This is a new variable. Jeffrey Williamson did an excellent work in the compilation of real wages for Third World countries across the world, for the period 1820-1940 (Williamson, 1999, 2000a, and 2000b). The nominal wages are weighted by the Cost of Living Index, accessible from the same source. The data appendices are frequently revised and are available on-line at Williamson's webpage. The latest available version of Williamson's wages has been reported, for those case where it is available. This source has, however, two limitations: Firstly, a good part of the countries in our sample are not included in the study. Secondly, the wages are not separated by sectors, so we would have to make assumptions about the peasant/manufacturer ratio if we want two separated wage series for industrial workers and agricultural labourers. A new variable "wages" reports the average wage in that country at that point in time, regardless of in which sector it has been originated. This is why we may want to maintain as well the Adelman and Morris wage related variables, because they are an indicator of differentiated evolution in industrial and agricultural wages. Actually, Mitchell's data on wages, which –although not available for all countries– have also the advantage of differentiating between agricultural and industrial wages (Mitchell 2003a, 2003b, and 2003c). Unfortunately, one cannot re-construct sector-separated wage series for many countries in the sample. This is why up-to-date sector-separated wages by Mitchell are compiled in the new database, but not used for data analysis directly. An index on the level of real wages country by country *à la* Williamson has been used in the calculations. Still, Mitchell's wage series have come to be very helpful in complementing the real

wage series available from Williamson. Because Williamson's wages are available for a reduced number of countries only, for most of the countries, wage estimates have been derived from Mitchell's wages by sector, taking into account the percentage of workers in every sector. Since these are money wages, they have to be converted into real wages. For this purpose, the Consumer Price Index (CPI) from the same source (Mitchell 2003a, 2003b, and 2003c) has been used for consistency.

Direction of Change in Average Real Wages in Industry

Complications in the calculation of homogeneous wage rates led to Adelman and Morris to classify real wages according to the *probable* direction of average change in every 20 years period, and whether the variation was weak or strong. Furthermore, they penalised volatility. These distinctions led to distinguish between five categories:

- strong upward movement
- upward movement, but not strong
- stable
- fluctuating with no trend
- downward movement

Incomplete series were an additional difficulty in the calculation of industrial wages. Incompleteness seems to be coming from partial reporting: either just one or two categories of industrial workers had their wage reported or it was reported for a brief period only (Morris and Adelman, 1988:231). It seems clear that a replacement of this variable by a historical real wage series would be beneficial. Up-to-date industrial wage series have been compiled in the new database, from Mitchell (2003a, 2003b, and 2003c), but there is not enough information to replace the Adelman and Morris proposal with an actual rate of change of real wage in industry.

Direction of Change in Average Real Wages or Income of Employed Agricultural Poor

The procedure to classify country-period observations (for instance, Argentina 1850-1870) into different categories was exactly the same to that described above for real wages in industry (see 'direction of change in average real wages in industry); thus leading to exactly the same possible categories in Morris and Adelman (1988). An

additional complication, specific of agricultural wages, was the seasonality of agricultural labour. Ideally, agricultural wage changes would have been substituted by the actual rate of change of real wage estimates according to nowadays historical statistics. Up-to-date industrial wage series have been compiled in the new database, from Mitchell (2003a, 2003b, and 2003c), but, unfortunately, only a few countries have enough information to replace the Adelman and Morris proposal with an actual rate of change of real wages in agriculture.

Demographic variables

Demographic variables capture movements of people, both the capacity of a society to grow in size as well as migratory moves. Three demographic variables summarise the data referred to population size and migration in the original database. This number has now been extended to four:

- total population
- population growth, lagged
- net immigration
- rate of net migration to total population growth

Total Population

Maddison (2003) compiles statistics of population after 1950 from the International Programs Center, US Bureau of the Census, October 2002 (USBC www.census.gov), except for China, India, and Indonesia; (the latter not in our sample). For the absent countries and estimates before 1950 he uses specific country sources detailed in Maddison (2003). The data available from the United Nations Population Division (UNDP) would have been the alternative source for population statistics. Maddison prefers the US Bureau of Census data over the UNPD ones because the former gives a more detailed account of short-term shocks. This is due to the fact that the United Nations interpolates between censuses. Therefore, USBC data capture ups and downs much better, while the UNPD data seems more suitable for studying long-run trends. We stick to Maddison's population series. In particular, the August 2007 updated version of the data has been used (accessible on-line at http://www.ggdc.net/maddison/).

Population Growth, lagged

The cumulative growth of the population for the past 20 years has been derived from the variable above (total population) and expressed in percental terms. This is an improvement over the original population growth reported by Morris and Adelman (1988), since they based their estimates of population in unofficial sources in many cases.

Net Immigration

What Adelman and Morris called "net immigration" actually reports the rate of net migration to total population growth (following variable). They never reported levels of net immigration in their final tables. However, they constructed a table showing the information available from different sources regarding migration levels (Morris and Adelman, 1988: Table A32). Statistics on migrations were extracted mostly from Ferenczi (1929), Ashworth (1962), and Mitchell (1976). Unfortunately, they never made a decision on which sources to use in order to give a point estimate for net migration. They limited themselves to place countries in one of 6 categories. So, in a sense, this table is unfinished. In the light of the information they compiled, there is scope for providing point estimates of net migration. Different sources differ in periodisation and amounts, even for the same period; therefore, there is a task of addition for different sub-periods and selection of sources. When information from different sources does not coincide, the most recent source has been used. No further alterations have been made to this variable, since still nowadays statistics available on migration are fairly incomplete. Consequently, we would not expect substantial improvement. Emigration is reported with a negative sign.

Rate of Net Migration to Total Population Growth

The rate of net migration to total population growth captures the direction and importance of migration compared to total population changes (Morris and Adelman, 1988:232). The original Adelman and Morris database does not have a variable under such name; they called it "net immigration", which now corresponds to a completely renovated variable. In the original database, six categories describe the importance of

migration for changes in population, ranging from the cases where more than one-third of the increase in population was due to *immigrants* (European settler economies) to the cases where more than one-third of the population change was due to *emigrants* (some European countries during the last third of the nineteenth century).

This variable can be improved by reporting the exact estimated percentages instead of grouping the observations in six arbitrary categories. The percentages of migration over population change have been calculated from the figures on migratory movements summarised in Morris and Adelman (1988: Table A32), together with the population variables described above.

Human Capital

The stock of human capital in the époque is represented by literacy, while the growth of human capital is represented by the spread of primary education. These two concepts (stock and flow) give rise to two variables:

- Adult illiteracy
- Rate of spread of primary education, lagged

One extra variable has been added to the database, namely

- Quality of education

Adult illiteracy

Adelman and Morris constructed a classification of countries according to the percentage of illiterate population. These percentages are based on specific-country sources. For this variable, we reverse the numeral re-codification. The variables range from A to J in the original, corresponding to 0 to 100 in the numeral scale. We have reversed this scaling to "from 100 to 0" in order to transform the illiteracy variable into literacy. Since we assign higher values to better outcomes, the transformation is intuitively more convenient.

Rate of spread of primary education, lagged

With respect to improvements on the primary education variable, the so called Barro-Lee Data Set gathers international measures of schooling years and schooling quality for 126 countries. Educational attainment is reported for different age groups and gender. Data are presented in five-year intervals for the period 1960-1990. Information in the Barro-Lee format is also available going back to the nineteenth century for most of the countries, consisting on university as well as primary and secondary school enrolment. Daniel Cohen presents a critic on the Barro and Lee dataset, and proposes a way to clean up the data. Further discussion on these data has been writen by Krueger and Lindahl, and Domenech. [3]

As far as historical data sources are concerned, Peter Flora provides a detailed description of the literacy rates by regions in Western Europe for years previous to 1960. This information has been constructed from the information available about army recruits and newly married people at the *époque*. This database goes back to 1815 and it is conveniently split into different cohorts and genders. In *State, Economy and Society in Western Europe 1815-1975* (vol I), Peter Flora provides the percentage of students enrolled in general, secondary and higher education, as a percentage of the total population.

Also Mitchell (2003a, 2003b, and 2003c) provides the number of children in schools, as well as the number of teachers. There are two advantages of using Mitchell's data instead of Flora's ones. First, the former contains data from all continents, while Flora refers exclusively to Western Europe. Second, Mitchell's publications are more recent and updated. While Flora published his collection of Western European educational statistics in 1983 and remains as it stood (Flora, 1983), Mitchell has kept renewing and updating his collection of statistics, as well as adding more countries. For these reasons, Mitchell is the preferred source. The number of students in primary education is recorded. However, this variable cannot substitute the original one, because for many

[3] There is a lot of criticism on the Barro and Lee dataset, concerning the adequacy of the way to measure human capital and the lack of quality of the data. It is worthwhile to explore those criticisms, starting by those by Krueger and Lindahl (2000) and De la Fuente and Doménech (2000). Also of interest: Cohen and Soto (2001), who propose a way to polish the data.

countries in the sample reporting does not start until 1870 and, therefore, rates of growth cannot be calculated for the period 1850-1870.

Quality of Schooling

Morris and Adelman (1988) acknowledge they could not include the quality of education, though they do not mention the reasons. Nevertheless, it stands quite clear that the measurement of the quality of education is nothing but trivial. In fact, the issue remains unresolved in the literature. There is a lot of criticism on the Barro and Lee dataset, concerning the adequacy of the way to measure human capital and the lack of quality of the data. See Krueger and Lindahl (2000), De la Fuente and Doménech (2000), and Cohen and Soto (2001). Below, we propose an indicator for the quality of schooling.

On the one hand, we have the number of pupils in primary education (in thousands). This variable has been extracted from Mitchell's international historical statistics. On the other hand, we have the number of teachers in primary education (in thousands), from the same source. The ratio of the two variables is a new variable representing the number pupils per teacher. In order to make it increasing with the quality of education, it is more intuitive to place the number of teacher in the numerator and the number of pupils in the denominator of the ratio.

A new variable (teachers per pupil) has been created by dividing the number of teachers by the number of students. The number of teachers per pupil intends to be an indicator of the quality of education, therefore adding the already noticed by Adelman and Morris missing aspect in the educational variables.

After compiling information for the number of students in primary education and the number of teachers for all possible countries in the sample, statistical analysis reveals that these two variables present a Pearson correlation of 0.953. i.e. the number of students and the number of teachers grew in the same proportion. Therefore the quality of education as proposed above was held practically constant throughout (on average). For this reason, it does not seem worrying that the amount of countries for which the new variable "quality of education" is available for the desired periods is rather limited. Because only about half of the sample is available, the variable "quality of education"

has to be excluded from the principal components analysis, which requires full sample.

Socioinstitutional influences

Socio-institutional variables in the database are the following:

- Urbanisation
- Attitudes toward entrepreneurship

Urbanisation

The advantages of urbanisation for social interaction are numerous, amongst which we can mention: markets, higher mobility, or political participation. The variable "urbanisation" measures the proportion of the population living in cities of more than 10,000 inhabitants. The exact percentage is not available in many instances; so, in this case, the Morris and Adelman (1988) classification scheme is convenient. The classification scheme of is as follows, in percentage of population living in towns of more than 10,000 inhabitants:

30% or more
between 20% and 30%
between 10% and 20%
less than 10%

Most of the information comes from Weber (1967).

Attitudes toward Entrepreneurship

The social attitude toward entrepreneurship refers to whether or not the social elite regarded favourably or not entrepreneurial success. This is a purely qualitative variable; so it is not based on statistics, but on informed judgments. Adelman and Morris took informed judgements from the literature and constructed this variable by means of contrasting attitudes in different countries. They divided the spectrum of social recognition of economic activities into 4 levels of social recognition, with which they created six categories, each one identified by a series of social signals:

- Category A: Social recognition was indicated by official recognition, intermarriage, and retention of successful industrial entrepreneurs in industry.

- Category C: Social recognition of entrepreneurial activity went to those who were already part of the elite.
- Category E: Landowners and primary product exporters gained social recognition, but not mere accumulators of mercantile wealth.
- Category F: Not common social recognition by the elite.

In categories B and D were some part of the country fitted in the previous category and some other part fitted in the following one.

The judgements made in the multiple sources consulted are difficult to refute as a whole, as well as it is difficult to reproduce or refute the classification made by Adelman and Morris. Consequently, no changes have been introduced in this variable, apart from numerical re-codification.

AGRICULTURAL INSTITUTIONS

Agricultural institutions refer to land tenure, land concentration, and, in general, institutional variables that may end up affecting land productivity. Innumerable land reforms have been carried out in many countries, though most failed in a way or another. Some, however, did succeed and effectively changed the holding traditions and plot structure. The following variables summarise institutional aspects of agriculture:

- Predominant form of land tenure and holding
- Concentration of landholdings
- Favourableness of land system to adoption of improvements

Since all three variables are qualitative, we follow the original Adelman and Morris classification schemes for all of them.

Predominant Form of Land Tenure and Holding

There is a single criterion for classification in this variable, and it is clearly stated: Systems of land tenure were classified according to the degree of control of the direct cultivator. This ranking principle is the one proposed by Boserup (1963). We find countries with predominant independent cultivators with full-right ownership at the top

of the seven categories (for instance, Scandinavian countries). Centrally controlled agriculture with some form of servile labour appears at the bottom (Brazil, Egypt, and Russia around mid-nineteen century). The entire classification scheme runs as follows, from top to bottom:

Independent cultivators, with full-right ownership

Idem, with some feudal obligations (taxes, kind or services)

Cultivators paying fix cash rent

Short-term tenants

Peasants with their own land, but subject to communal controls

Centrally managed land, usually not by the owner, with hired workers

Centrally managed land, usually not by the owner, with some form of servile labour

Sources for classification are mostly country-specific literature, which difficults a revision of this variable.

Concentration of Landholdings

The original database has seven classification categories for the concentration of landholdings, based on the size of holdings. On the top category, the top 10% of landholders have at least 75% of the cultivable land. On the bottom categories, either holdings are very small, shared, or the predominant form is communal land. A simplified version of the classification is as follows:

The top 10% of landholders have at least 75% of the cultivable land

Extreme concentration of landholdings, but not as much as the first category

Large holdings

Middle-size holdings

Small holdings

Holdings are very small, or shared

The predominant form is communal land

Also other secondary considerations enter the classification, such as, the form of labour. For full details, see Morris and Adelman (1988), table A38:413-416.

There are two aspects to have into account when studying concentration of landholdings:

First, renewed tenure schemes and plot concentration has a direct effect on productivity, as shown in case studies below, but the direction of the effect does not always coincide. Land concentration raises productivity by means of scale effects; while some types of innovation spread more easily within a small plot structure. Moreover, the debate between hired versus family labour has also revealed some contradictory outcomes depending on the case study. Morris and Adelman pose some illustrative examples on this complexity in the following passage:

> 'The relationships between land systems and productivity were complex: biological innovations often spread on small family holdings (Hayami and Ruttan 1971); mechanized techniques spread first to larger farms (Jensen [Denmark] 1937; Clapham [Germany] 1936). Independent family farms provided strong work incentives and cheap labor, but their limited assets made investment risky (Wright [United States] 1978). Larger-scale capitalist owners with wage labor could not always compete (Smith [Japan] 1959). However, in England, capitalist tenants with hired labor raised productivity significantly (Ernle [Great Britain] 1936). In contrast with these cases, large-scale estate owners with aristocratic values accumulated wealth in land and slaves and were slow to adopt new techniques (Dovring 1965)' (Morris and Adelman, 1988:45).

Thus, we cannot generalise. The overall *direct* effect of agricultural institutions on productivity is ambiguous and/or strongly depends on the case.

and

Second, renewed tenure schemes and plot concentration changes social relations in the countryside. Therefore it has an effect on social capital. Actually, the first social transformation comes from the economic system. As a matter of fact, traditional communal practices are lost with capitalism. This will eventually be reflected in the change in the agricultural institutions, which are forced to adapt to the economic environment. We will use the change in institutions as a reflection of the loss in communal behaviour, so traditional in agricultural practices. Communal practices were common in most parts of the world, especially before the marketing and export oriented exploitation. (for evidence see Furnival, 1931; Indian examples might be found in Mukerjee, 1972, among others). Non-capitalised agriculture (self-consumption oriented)

tends to be more oriented to the benefit of the community, being this a group of families or neighbouring farms. Barter instead of monetary transactions potentiates gifts, communal labour, and sharing of yields, expecting some reciprocity. This tendency will progressively shrink as non-agricultural population increases. We would expect the most modern institutions to go in detriment of the common good tasks or attitudes. So, progress in agricultural institutions works counter-intuitively.

Currently, higher punctuation is assigned to higher concentration of land. For the reasons stated above, there could be an argument for reversal of punctuation. However, the effect of land concentration on productivity is still ambiguous.

Favourableness of Land System to Adoption of Improvements

This variable is a combination of the previous two; weights for every country according to *a priori* judgement of the authors of the database. The specific weights were never given, so we have to accept their classification scheme. This time, they were able to differentiate up to nine different categories for classification. These are, from top to bottom:

Independent cultivators without extreme concentration of landholdings

Tenants paying fix cash rents, without extreme concentration of landholdings

Family farms

Tenants paying fix cash rents on family-size holdings

Centrally managed, with hired workers

Peasants with extremely small holdings

Short-term tenants

Peasants with some form of communal control

Large states with some form of servile labour

This long categorisation is remarkably detailed, and worth profiting from. It is difficult to find a qualitative variable like this one, with such a detailed classification. We stick to the suggestion by Adelman and Morris.

POLITICAL INSTITUTIONS

Social and political structures go hand in hand. Their relationship has been traced back as far as the Middle Ages. Perry Anderson's thesis (1974) is supported by Moore (1966) and Warren (1980): 'Anderson (1974) emphasizes the importance to Europe of the feudal legacy of fragmented political authority. (...) Dispersal of sovereignty permitted a great diversity of populations and languages to persist. According to Anderson, a similar dispersion of absolute power within feudalism occurred in Japan' (as reported by Morris and Adelman, 1988:47). This division of authority promoted a 'vigorous and independent class of town dwellers' (Moore, 1966:148), and 'mutual engagement derived from feudal relations of vassalage' (Morris and Adelman, 1988:47). This exemplifies how political systems may influence social structures.

At the same time, new social classes had some influence on political structures' transformation. Anderson and Anderson (1967) argue that the post-industrial change in European political institutions is conducted by social pressures emerging from a transformation of the society, such as the emergence of the middle class. Quoting Morris and Adelman, 'According to Anderson and Anderson (1967), industrialization and concomitant social changes in Europe induced political change. The growing middle class sought participation in political affairs, educational opportunities, and social advance for themselves and their children' (Morris and Adelman, 1988:46). Karl Deutsch does not speak of social capital, though he speaks of 'social mobilisation'. 'Deutsch (1953) argues that "social mobilization" (the breakdown of old commitments, making people available for new ones) pressured political practices and institutions to change' (*ibid.*).

The examples above suggest that society and politics are interrelated. The World Bank has been producing aggregate good governance indicators for the last decade, starting 1996. This is known as the Governance Matters database, which counts with the following indicators: voice and accountability, political stability, government effectiveness, regulatory quality, rule of law, and control of corruption. However, this is a relatively new project and, thus, does not count with past data indicators (see Kaufmann, Kraay, and Mastruzzi, 2007 for the last round, with data starting in 1996).

The original Adelman and Morris database has a good collection of political indicators for the nineteenth century. These are:

- Domestic economic role of government
- Character of political leadership
- Strength of national representative institutions
- Political stability
- Foreign economic dependence
- Colonial status

To these, some additions arising from recent research have been incorporated, namely the polity index coming from the latest version of the Polity IV database, and the updated point estimates corresponding to government expenditure relative to GDP. On top of older research such as that of Banks and Textor (1963) or Banks (1971), other recent efforts such as that of Alesina and Perotti (1996) Sociopolitical Instability Index, Heinsz (2000) Political Constraints Index, and Roe and Siegel (2007) 30 year moving average of one of Banks (2005) indicators have been evaluated.

Domestic Economic Role of Government

Morris and Adelman (1988) ranked all countries into five categories, according to the extent to which government expenditure contributed to domestic economic activity. Given that government expenditure figures are available, it seems an improvement to switch to **public expenditure relative to GDP**. Mitchell (2003a, 2003b, and 2003c) provides total central government expenditure, which can then be re-calculated as a proportion of GDP. A disadvantage of this strategy is that it does not differentiate between domestic and external public expenditure, as the Adelman and Morris indicator did. Nevertheless, the latter, departing from only five different categories, can really be made much more precise by taking public expenditure statistics. In any case, the objective of the database is different from that intended by Adelman and Morris. While they were interested into the effect of government expenditure into economic activity, public expenditure relative to GDP may be for us an indicator of how much citizens are willing to entrust to the government.

Character of Political Leadership

This variable features the socio-economic character of political leadership. Morris and Adelman (1988) identified four different categories for this variable, depending on which groups had power in the national government. These are, from top to bottom category:

Indigenous had the power, including direct representation in national parliaments

Indigenous and elite shared the power, including direct representation

National traditional elite had the power, but rising influence of new entrepreneurs

National traditional or colonial elite had the power

The classification was based on descriptive information extracted from political and economic country-specific histories. There exist other classifications of the character of the governors, but the Morris and Adelman (1988) classification is quite unique, in the sense that no other alternative captures the same concept, relating elite to social origin.

With a technique adopted from the cross-cultural analysis, Banks and Textor created pair wise comparisons of 57 polity variables, referring to practically all countries in the world (Banks and Textor, 1963). Their book is published in 1963 and the data make reference to the époque. However, their political modernisation variable refers a historical event: the influence of colonization.

The potentially relevant variable in Banks and Textor is the "political modernisation" variable. The classification is by historical type (Banks and Textor, 1963:77-88):

A. Early European or early European derived (early modernizing European society or offshoot)
B. Later European or later European derived (later modernizing European society or offshoot)
C. Non/European autochthonous (self-modernizing extra-European society).
D. Developed tutelary (developed society modernizing under tutelage).
E. Undeveloped tutelary (undeveloped society modernizing under tutelage).

Nevertheless, as one-shot historical event, it does not have any dynamic aspect, or process of change involved, as in the spirit of Adelman and Morris. The authors

acknowledge this handicap and shelter in the difficulty in doing so (Banks and Textor, 1963:7).

In later work, Banks (1971) compiles data describing politics, economy, population and communications starting as early as 1815 in some cases and running up until 1966. The type of data collected is similar in nature to the compilations by Mitchell discussed above, but different in sourcing. The main source for Banks was *The Statesman's Yearbook*, a yearly publication which offers an overview of the state of affairs in terms of political regime and economic situation of the economy for every country in the world since 1864. Banks and his team systematically compiled data from *The Statesman's Yearbook* and additional sources and published the data tables as the *Cross-Polity Time-Series Data* in 1971 (Banks, 1971). The potentially relevant variables in Banks (1971) and its classification are as follows:

Type of Regime

(1) Civilian

(2) Military- Civilian

(3) Military

(4) Other

Head of State

(1) Monarch

(2) President

(3) Military

(4) Other

Effective executive (Type) - Same typology as Head of State; who exercises primarily influence in shaping the decisions of the nation.

These measures are far too general and in no way related to the social character of the leadership. No inspected alternative is superior to Morris and Adelman (1988) in terms of capturing social integration.

Finally, the Polity IV project developed by the Center for International Development and Conflict Management (CIDCM) at the George Mason University and the University of Maryland does an excellent job. Subsequent versions of this international historical database have been constantly revised, and the fourth version contains records of basic

institutional characteristics from 1800 to 2004, on a yearly basis. Actually, the most interesting candidate in the polity database, is no longer included in recent version, and one has to go back to the second version, Polity II, in order to find the "scope of government action" variable. This variable 'is a measure of the extent to which the state uses its powers of regulation and command to direct social and economic activity' (Marshall and Jaggers, 2005:35). Although this variable is not exactly the same as Adelman and Morris tried to capture, it goes in the same direction, relating the authorities to the level of involvement in the society. 'It is a continuum which refers to the extent to which all levels of government combined –national, regional, and local– attempt to regulate and organize the economic and social life of the citizens and subjects of the state' (*op. cit.*: 37). Nevertheless, and despite the broad coverage of this variable, the scope of government action as in the polity database has been excluded from the principal components analysis due to poor variability of the data. The original variable socio-economic character of political leadership, also satisfying and conceptually very close, has been used instead.

Strength of National Representative Institutions

This variable in the database is an index constructed from 3 indicators (all information from country-specific political histories):

1. The extent of legislative power of the parliament - In the nineteenth century, the existence and strength of a parliament was a strong indication for separation of powers.
2. The extent of popular suffrage, which is the most direct indicator of representativeness of the government.
3. The weight of opposition parties.

1 and 2 were strongly positively correlated in the nineteenth century. Countries with separation of powers tended to require a more accessible minimum of property qualifications to become a voter. Adelman and Morris did not make use of the extensive international list of political characteristics in Banks (1971) at any point in the construction of this variable. They made use of individual country histories instead (Morris and Adelman, 1988:237 and Table A44:441-6). However, Banks (1971) international political variables are worth exploring. Banks (1971) potentially relevant variables and its classification are the following:

Effective executive (selection)

(1) Direct election

(2) Indirect election

(3) Nonelective

Parliamentary responsibility

(0) irrelevant

(1) absent

(2) incomplete

(3) complete

Legislative effectiveness

(0) none

(1) ineffective

(2) partially effective

(3) effective

Legislative selection

(0) none

(1) nonelective

(2) elective

Legislative elections (number per year)

Banks' database constitutes quite a good effort; it is a systematic account of general political features. But, firstly, it is not a significant improvement over Morris and Adelman (1988) in terms of content. The main difference between the two is sourcing: While Adelman and Morris surveyed country histories, Banks systematically scrutinises *The Statesmans' Yearbook.* Secondly, there are more precise representativeness indicators arising from more recent research.

Nowadays, it is common to find the use of a simple democracy dummy as an indicator of the representativeness of the government. Neverthesless, the Adelman and Morris indicator is no doubt more sophisticated. The original variable in the database is broader in content and more detailed in range of categories.

At this respect, the Polity IV project records openness of political institutions in a more precise 0 to 10 democracy score. The polity data archive incorporates some of the variables created by Banks. This is preferable to a simple democracy dummy and also to

Banks (1971) because it is more precise. With comparison to Morris and Adelman (1988), the Polity IV democracy score is narrower in content, representing a very particular aspect. However, the Polity IV data archive combines the democracy score with an autocracy score (closeness of political institutions) with the same range of variation, generating a -10 to 10 scale called polity variable. A variation of this variable called POLITY2 is preferred, because it assigns normal range polity scores to observations corresponding exceptional situations. A disadvantage of using this alternative is that periods of interruption such as foreign occupancy due to war are left blank; and we need full record for the principal components analysis. Luckily enough, no observations of interest fall under that category.

Political Stability

The Adelman and Morris indicator for political stability encompasses three different aspects in a single index. These three aspects are the following:

- Continuity in national political decision making
- Intense non-violent political strife
- Domestic violence

Adelman and Morris combined these three features into four categories, which describe their political stability indicator. The four categories are the following:

	Continuity	No strife	No violence
Category A	yes	yes	yes
Category B	no	some	yes
Category C	no	some	no
Category D	no	no	no

All information necessary was compiled from political histories.

Staying aside from the nowadays popular political instability indicators, the Adelman and Morris variable does not include democracy in their indicator; this leaves open the possibility of treating democracy as a singular differentiated feature, and it is indeed treated as a separate variable. Representativeness of the government is included in the former variable (strength of national representative institutions), which is, on the other hand, a broader concept than democracy alone. Still, we may want to do a brief survey

of what sort of political stability indicators are the most common in current research.

In current research, the most widely used indicator for political stability is the one introduced by Alesina and Perotti (1996). This index is called the Socio-political Instability index, and comprises political and mass assassinations, coup attempts (whether successful or not) and a democracy dummy. The time span of their indicator runs from 1960 to 1982. As Roe and Siegel point out, "this has been a widely used and respected measure of political instability" (Roe and Siegel ,2007:16). One reason why this index is not used here is that it is only available for recent decades. In addition to the time span constraints, this index is not necessarily more adequate than the original one devised by Adelman and Morris. In comparison, they are very similar in content, the only significant difference being the use of a democracy dummy by Alesina and Perotti (1996) instead of a continuity in policy making indicator in the index devised by Morris and Adelman (1988), as one of the 3 aspects considered. The latter might seem even more adequate, especially given the fact that democracy has already been taken into account in the former variable "strength of national representative institutions".

An early starting collection of data is that pioneered by Arthur Banks. Some of his political indicators start as early as 1815, and could potentially be relevant here. His book *Cross-Polity Time-Series Data* compiles –among other things– political indicators (Banks, 1971). Some of the materials overlap with the Adelman and Morris database. The latest version is an on-line resource, the *Cross-National Time-Series Data Archive* (Banks, 2005). A 30 year moving average of one of the variables in this data archive was used by Roe and Siegel (2007) as an indicator of political instability. The variable in question is named government crises, originally available since 1919 from Banks (2005); it refers to "any rapidly developing situation that threatens to bring the downfall of the present regime" (Banks, 2005, http://www.databanks.sitehosting.net/, list of variables). This variable is conceptually interesting, but it starts after World War I, just after the period under study. Likewise happens with a relevant variable in an earlier version of his work: variable 27 in the aforementioned Banks and Textor (1963) is "governmental stability". It is classified making historical reference to the First and Second World Wars; so it is not possible to apply Banks and Textor classification to the present study.

On the other hand, the continuously updated Polity IV Project offers the DURABLE variable, which provides the age of regime, in years. This is a very simple measure of political stability, which is a desirable characteristic. However, young regimes do not necessarily have to be destabilising. This is especially true in the case of colonies earning some degree of independence, like transfer of powers, or cases of formation of nation-states, especially common in the period under study. So, the DURABLE variable is not the preferred one. The Adelman and Morrris political stability variable encompasses several aspects directly referring to society, in particular referring to civil unrest, which are especially relevant when getting a picture of the quality of society. Thus, the latter is preferred.

Finally, in recent literature, there is a completely new alternative arising from the work of Witold Heinsz. According to Heinsz and Zelner (Measures of Political Risk), accounting for democracy and political stability is not enough to create the right atmosphere to attract investors in a country, 'but rather the ability of the government to craft a credible commitment to an existing policy regime' (Heinsz and Zelner, *Measures of Political Risk*, online resource:4). So, the trust of investors in a country depends on what investors perceive or expect, rather than on what investors actually observe. With this in mind, they propose 'a measure that is objectively derived with the explicit goal of measuring the likelihood of changes in the policy regime' (*ibid.*). This is known as the Political Constraints Index (POLCON) and rates from 0 to 1 the extent to which a political actor is constraint in their decision making. The procedure is explained in detail in Heinsz (2000). The ultimate objective of this index turns out to be exactly the same as the political instability variable devised by Adelman and Morris, which encompasses information on civil and political tensions as well as political turnover (Morris and Adelman, 1988:237).

There are two reasons to prefer keeping the original variable in the Adelman and Morris database instead of the more sophisticated Heinsz index here. Firstly, the social character of the Adelman and Morris variable against the strictly political Heinsz variable is more appropriate to the context of capturing social change. Secondly, despite the extraordinary work of Heinsz in compiling data starting as early as 1815 for some cases, his index is still limited in the number of years available for the period required. So, a newly generated variable necessarily filling the blanks due to methodological requirements would forcefully have to rely on far too many assumptions.

Foreign Economic Dependence

The degree of foreign economic dependence reported by Morris and Adelman (1988) is an aggregate measure of seven indicators, all of which are can take the value "heavy", "moderate", or "modest". The seven indicators are the following:

- Primary exports dominate the export sector
- Foreign ownership and control of factory industry
- Importance of expatriate and government economic initiatives
- Foreigners' control of trade, distribution, and related financial services
- Dependence of expanding sectors on specialised foreign skills
- Dependence of the government on foreign loans
- Foreign capital financed domestic investment

(See Morris and Adelman, 1988: 156-7).

With the answer to this seven indicators, Adelman and Morris could classify observations into one of seven categories:

Heavy dependence in all indicators

Heavily dependent production structures

Moderately dependent production structures

Moderate dependence in average

Little dependence

Independent production structures; only with countries at the same level for the rest

No dependency features

There are no specific weights given to the different aspects of economic dependency. However, status of dependence of production structures was prioritised over the rest. We inverted the numerical re-codification, in order to assign higher punctuations to economic independence, unlike the original. This transformation makes the variable more intuitive.

Colonial Status

Colonial status is another qualitative variable that was categorised by Morris and Adelman (1988). The six categories in this case run from being a colony to being a

colonial power. These are:

Colonies run by expatriates

Colonies run by settlers

Political independent but under strong colonial power

Neutral

Modest colonial holdings

Major colonial holdings

The source used for classification was the Encyclopaedia Britannica.

MARKET INSTITUTIONS

How much did the market institutions evolve in this period? Adelman and Morris did extensive previous work on the measurement of market institutions. Earlier to the publication of the database we are working with, they prepared twelve classificatory indicators for the level and spread of market institutions (Adelman and Morris, 1978). Morris and Adelman (1988) summarises previous work in this area in only three variables:

- Development of market institutions
- Spread of market institutions
- Spread of market institutions, lagged

Development of Market Institutions

The nineteenth century is characterised among other things by the development of markets. The variables under the name of market institutions summarise the development of the market system. Examples of nineteenth century market developments are illustrated in the following passage:

> 'Historical case studies suggest several (...) features of market systems in the nineteenth century. Formal, settled markets replaced migratory or occasional ones. Specialized institutions evolved for commodity exchange and for mobilizing labor, land, and capital –for example, staple goods exchanges, joint stock companies, and land mortgage and investment banks. Finally, market institutions expanded the geographical area of their activities, reaching

> nationwide scope in countries with more developed systems'. However, 'The spread of market institutions (...) was uneven outside Europe and the United States. In some areas, national institutions evolved for only a limited range of export or government-subsidized goods. In others, colonial powers dominated export marketing systems while domestic institutions for goods and factors remained rudimentary' (Morris and Adelman, 1988:63).

Polanyi (1944) underlines that social changes are a pre-requisite for markets. 'Polanyi saw labor as "embedded" in a series of quite "natural" social relationships that made the construction of market institutions and impersonal exchange extremely difficult. However, the advance of capitalism and the commodification of labor created "disembedded" markets' (Blyth, 2002:3). So, before formal markets, natural social relationships were a necessary pre-requisite for exchange. The arrival of the market facilitated exchange. Mark Blyth has recently reinterpreted Polanyi's thesis on twentieth century institutional transformations. He argues that whether we operate under the old or the new institutions does not make such a big difference in term of embeddedness (Blyth, 2002:4). This idea had been previously underlined by Ruggie (1982), Helleiner (1994), McNamara (1998), and Kirshner (1999). All of them termed Polanyi's description of the 'great transformation' 'embedded liberalism'. Translated into our terminology, this means that the view that market institutions do need social capital to work is recaptured.

Morris and Adelman (1988) provide a composite indicator, which is the result of a single-principal component analysis including the following market indicators:

- Level of development of domestic commodity markets
- Level of development of domestic labour markets
- Level of development of domestic capital markets
- Level of development of domestic land markets

These indicators are based on previous publications (Adelman and Morris, 1978). We have not used classification schemes for the level of development of market institutions. Instead, we directly used the scores of the composite index.

One observation should be made about the suitability of the principal component analysis used to build the composite indicator:

> 'Developments within different market types proved interconnected. The growth of national goods markets coincided with widespread wage labor, specialized capital institutions, and monetized land transactions. The dualistic development of commodity markets was associated with dualistic factor market expansion (Morris and Adelman, 1988:67).

So, market developments move together. Development in one sector or region stimulates development in others; therefore, there will be no problems of trade off in the development of markets indicator.

Spread of Market Institutions

The rate of spread of market institutions is a composite index derived with the same technique as the level of development of market institutions (previous variable). The single indicators used for this composite index were the following:

- Rate of spread of domestic commodity markets
- Rate of spread of domestic labour markets
- Rate of spread of domestic capital markets
- Rate of spread of domestic land markets

Details for these indicators can be found in Adelman and Morris (1978). The four single markets produce a single-principal component analysis, which scores are the compound index. As in the previous variable, we skip the classification and take instead directly the compound index scores.

Spread of Market Institutions, lagged

This variable is simply a lagged variable of the previous one.

2. IV. OVERVIEW SUMMARY

This chapter is a review of the Adelman and Morris database for the nineteenth century. Its particularities and adequacy for the project are discussed. Some useful comments made by other authors are also compiled and examined.

It has been possible to improve the original Morris and Adelman (1988) database in several ways. First, these data were not available in nowadays-readable electronic format. They should have been digitalised. Second, the original authors used a rather peculiar codification of the variables. They have been recodified in a format more convenient for statistical analysis. Third, Morris and Adelman (1988) data appendix provided some extra variables not listed in their final tables. These extra variables were all alternatives to the ones eventually included in the final tables, and therefore, should be (and have been) considered. And fourth, a vast amount of new material has appeared since the publication of the original database in 1988. There was enough new material to update it with the new information coming from various more recent data sources. Overall, a completely renewed database arises.

APPENDIX 2.A

LIST OF COUNTRIES

Recoding of variable 'country' into 'countryn' in order to assign a value to every country:

COUNTRY Old Value	COUNTRYN New Value	Value Label
Argentina	1	Argentina
Australia	2	Australia
Belgium	3	Belgium
Brazil	4	Brazil
Burma	5	Burma
Canada	6	Canada
China	7	China
Denmark	8	Denmark
Egypt	9	Egypt
France	10	France
Germany	11	Germany
India	12	India
Italy	13	Italy
Japan	14	Japan
Netherlands	15	Netherlands
New Zealand	16	New Zealand
Norway	17	Norway
Russia	18	Russia
Spain	19	Spain
Sweden	20	Sweden
Switzerland	21	Switzerland
UK	22	UK
US	23	US

APPENDIX 2.B

LIST OF VARIABLES

This is the list of variables in the database, in the same thematic order that has been used above in the main text of the chapter.

There are 73 variables in total in the new database, out of which 3 operate as strings (COUNTRY, YEAR, COUNTRYEAR) and 70 are numerical (all the rest), 35 of which are included in the principal components analysis. This is a list of the complete database, including operative variables and those not used for the principal components analysis, which are signed with a star *. Principal components analysis and results are explained in detail in the next chapter.

Strings:

COUNTRY = List of 23 countries above.

YEAR = For every country, the new database accommodates data for several time cuts: 1830, 1850, 1870, 1890, and 1910. 1870 and 1890 are complete; 1850 and 1910 are mostly complete, with just a few blanks; 1830 is has been added to keep record of the few variables available for that time cut, however, most are blanks.

COUNTRYEAR = This variable is a combination of country and year. It identifies observations.

Numerical variables are as follows:

*INCGROUP = Classification for level of per capita income. Re-codification: Enhanced precision scale 100 to 0 corresponding to range A to F in Morris and Adelman (1988). Point estimates (INCOME) used instead in the computations.

*INCOMEAM = Index of level of per capita income (UK 1890=100). Information summarised by Morris and Adelman in the variable INCGROUP. Point estimates (INCOME) used instead in the computations.

INCOME = GDP per capita as in Maddison (2003), updated with the latest version dated August 2007 (available from http://www.ggdc.net/maddison/).

Note: No blanks are allowed for the principal components analysis. As a result of this need for filling in the blanks, some missing data have been interpolated, and, in exceptional cases, extrapolated when needed at a constant annual growth rate.

*INCOMGRAM = Classification for rate of change in per capita income in the past 20 years. Re-codification: Enhanced precision scale 100 to 0 corresponding to A to D- in the original.

INCOMEGR = GDP per capita growth derived from INCOME.

INDUTECH = Level of development of techniques in industry. Re-codification: Enhanced precision scale 100 to 0 corresponding to A to F in the original.

INDTECGR = Classification for rate of improvement of techniques in industry (lagged, referred to the last 20 years). Re-codification: Enhanced precision scale 100 to 0 corresponding to A to G in the original database.

AGRITECH = Classification for level of development of techniques in agriculture. Re-codification: Enhanced precision scale 100 to 0 corresponding to A to G in the original.

AGRTECGR = Classification for rate of improvement of techniques in agriculture. Re-codification: Enhanced precision scale 100 to 0 corresponding to A to F in the original.

AGRILGRP = Classification for percentage of labour force in agriculture. Re-codification: Enhanced precision scale 100 to 0 corresponding to A to G in the original.

*AGRICLAB = Percentage of labour force in agriculture (calculated from Mitchell, 2003a, 2003b, and 2003c).

Notes: - Percentage of total population used when percentage of active population not available.

- Male and female labour force has been aggregated.

- Oscillations in actually reported years in many cases. Special cases: India and Italy 1890, with estimation actually corresponding to 1901. Russia and Switzerland 1890 is for 1900. China 1910 corresponds to 1930 figure. Russia 1910 is actually the figure corresponding to 1926.

- Morris and Adelman (1988: Table A9) used when no close estimate was available from Mitchell. From this table, most recent study used when discrepancies between sources.

*INDLAB = Percentage of labour force in industry (calculated from Mitchell, 2003a,

2003b, and 2003c). We are only going to make use of this variable as a weight in special cases in order to calculate wages.

Notes: - This is not a complete series, it contains only the observations we need for further computations.

- The amount of male population working in "commerce, finance, etc" and "transport & communications" in Sweden 1890 (necessary to calculate the percentage of people working in industry) has been interpolated from the 1880 and the 1900 figure, at a constant growth rate.

- The amount of people working in the manufacturing industries and construction in the USA 1830 (necessary to calculate the percentage of labour force in industry) has been interpolated with the figures corresponding to 1820 and 1840, at a constant growth rate.

*TOTALAND = Total land area in thousands of squared kilometres (Source: Clark, 1957: Table 33). Recovered from Morris and Adelman, 1988: Table A14, although it is not used in the calculations. FARMLAND is used instead because total land area does not take into account the varying quality of land, while farm land does.

*FARMLAND = Farmland in thousands of squared kilometres (Source: Clark, 1957: Table 33). Recovered from Morris and Adelman, 1988: Table A14. This variable is necessary for the calculation of POPXFARM.

POPXFARM = Population per square kilometres of farmland. Calculated from POP and FARMLAND. This variable corresponds to the "relative abundance of agricultural resources", as described in the main text.

INTRANSP = Level of development of inland transportation. Re-codification: Enhanced precision scale 100 to 0 corresponding to A to E in the original.

TRANSPGR = Classification scheme for rate of improvement of inland transportation (lagged). Re-codification: Enhanced precision scale 100 to 0 corresponding to A to E in the original.

*XGRGROUP = Classification scheme for rate of growth of total real exports. Re-codification: Enhanced precision scale 100 to 0 corresponding to A to D in the original. TRADE used instead.

*AVGXGR = Average annual rate of growth of real exports (%). Recovered from Morris and Adelman (1988: Table A18). TRADE used instead.

*AVGXGRP = Average annual rate of growth of exports in current prices (%). Recovered from Morris and Adelman (1988: Table A18). Real exports

classification preferred in the original; this allows comparisons.

*EXPORTS = Value of exports at constant prices (derived from Maddison, 2006: Table F-2).

Notes: -The case of Burma is very difficult because there are no data for this period. According to per capita GDP at constant prices, Burma's economy seems to have been completely stagnated over the period under study. There is no reason to think that exports would have behaved differently. During this period, per capita GDP at constant prices moved between 500 and 600 international dollars of 1990. That is about the same level Burma's economy oscillated between 1957 and 1973. So let us use exports data corresponding to this period in order to calculate the export to GDP ratio, about 250 international dollars of 1990.

*CURRENTX = Current aggregate external trade, in millions of local currency (from Mitchell, 2003a, 2003b, and 2003c). This variable is necessary to calculate relative trade (TRADE).

Notes: - For countries where current GDP exists for some years but not for all, the gaps have been filled by assuming the same rate of conversion from current to real GDP as the closest existing years present with Maddison's real GDP.

- Canada 1950 corresponds to 1967, since the latter is the first date available (but so is CURRENTGDP, so no mismatch in the calculation of TRADE); Argentina 1950 corresponds to 1964 for the same reason.

- Russia 1830 has been transformed to the new paper rubles for homogeneity with posterior figures at a rate of 4 to 1, as suggested by Mitchell (2003a:585).

- Switzerland 1870 has been extrapolated from 1885 (first year available) and 1890 at constant annual growth rate.

- Germany 1870 has been extrapolated from 1880 (first year available) and 1890 at a constant growth rate.

- Italy 1850 corresponds to 1861.

*CURRENTGDP = Current GDP, in millions of local currency (from Mitchell, 2003a, 2003b, and 2003c). This variable is necessary to calculate relative trade (TRADECURR and GOVREL).

Notes: -USA and Canada's figures correspond to GNP.

- Belgium, Germany, Netherlands, Spain, and Switzerland's figures correspond to net national product (NNP).

- Italy and Sweden 1850 has been extrapolated from 1961 (first year available)

and 1870 at a constant annual growth rate.

- Netherlands 1890 has been extrapolated from 1900 (first year available) and 1910 at a constant growth rate.

- Spain 1890 has been extrapolated from 1901 (first year available) and 1910 at a constant annual growth rate.

- Switzerland 1910 corresponds to the 1913 level.

*TRADECURR = Volume of exports relative to GDP, evaluated at current prices. CURRENTX/CURRENTGDP. TRADE preferred instead.

TRADE = Volume of exports relative to GDP, evaluated at constant prices. EXPORTS/INCOME. This variable measures the degree of internationalisation of the economy, substituting the rate of growth of exports.

Note: For the cases where exports data are not available nor can they be estimated safely, the Adelman and Morris scale for export growth group classification (XGRGROUP) is followed, assigning an average if there are different exports ratios falling under the same classification score. These countries are Egypt, Japan, New Zealand, Russia, and Sweden.

SHIFTX = Classification for degree of shift in structure of export sector. Re-codification: Enhanced precision scale from 100 to 0 replacing the original A to D.

INDWCHAN = Classification for direction of change in average real wages in industry. Re-codification: Enhanced precision scale from 100 to 0 corresponding to A to E in the original.

AGRWCHAN = Classification for direction of change in average real wages or income of the employed agricultural poor. Re-codification: Enhanced precision scale from 100 to 0 corresponding to A to E in the original.

*INDWCURR = Index of wages in industry (Mitchell, 2003a, 2003b, 2003c). Variable used to calculate INDW. 1900 = 100 like Williamson's most generalised base year for wages, for harmonisation purposes. Industrial wages include extractive industry, manufacturing industry, and construction. Sub-sectoral wages have been averaged out. Male and female wages have been aggregated.

Notes: - Except for the last observation corresponding to 1910, Australian data correspond to a simple average of wages in New South Wales and Victoria. The base year for the Australian series has been harmonised to 1900, departing from 1890 = 100 for all observations except the last one corresponding to 1910, for which the original base year was 1960 = 100.

- USA wage index has been homogenised to 1900 = 100, departing from 3 different base years for different sub-periods: 1830 = 100 until 1860, 1865 = 100 from 1860 until 1900, and 1955 =100 starting 1890 for some sub-sector onwards. The were 2 components available for the observation corresponding to 1890, the wage for all non-farming employees (with base year 1955) and the wage strictly for manufacturing (with base year 1955). These two have been transformed to a common base year 1900 = 100 and then averaged out.

- USA 1870 corresponds to an average of all non-farm employees.

- Belgium 1830 has been extrapolated from 1843 and 1850, following the steady tendency.

- Canada required 2 changes of base, one for 1890 data (from base year 1891 = 100) and one for 1910 data (from base year 1955 = 100), both to a common base year 1900 = 100.

*AGRWCURR = Wages in agriculture (Mitchell, 2003a, 2003b, 2003c). Variable used to calculate AGRW. 1900 = 100 like Williamson's most generalised base year for wages, for harmonisation purposes.

Note: - For Australia, there has been a change of base index in the money wage from 1955 = 100 to 1900.

- UK data correspond to England and Wales.

- USA required a change of base year from 1909 = 100 to year 1900.

*WHPRICES = Wholesale prices, base year 1913 (1914 in some cases) =100 (from Mitchell, 2003a, 2003b, 2003c).

Notes: - Change of base year for Japan from 1868 to 1913.

- New Zealand 1870 set up to 95, level similar to 1890. Prices do not show tendency to escalate (1910 level is even lower than 1890).

- France, change of base year from 1901-10 to 1913.

- UK 1830, change of base year from 1821-5 to 1913.

- Denmark, change of base year from 1891-1900 to 1913.

- Netherlands, change of base year from 1953 to 1913.

- Australia and Canada, change of base year from 1938 to 1913.

*CPI = Consumer Price Index 1913 (1914 in some cases) = 100 (Mitchell, 2003a, 2003b, 2003c).

Notes: - Belgium 1830 corresponds to 1835.

- France 1830 corresponds to 1840.

- Australia and Italy 1850 corresponds to 1861.

- Netherlands 1870 corresponds to 1880.

- Norway 1890 corresponds to 1901.

- UK suffers a change of base year from 1851 to 1914 in 1872. All UK price indices have been converted to 1914 base year.

- Canada 1890 corresponds to 1900.

- Argentina, Burma, Egypt, India, Japan and Spain 1910 correspond to 1913.

- Brazil 1910 corresponds to 1912.

- Australia and New Zealand appear with 1938 as base year in Mitchell (2003c:961). A change of base year has been performed (new base year 1913).

*PRICES = Consumer Price Index 1913 (1914 in some cases) = 100, substituted by wholesale prices when CPI not available for all necessary years (Mitchell, 2003a, 2003b, 2003c). Variable used to calculate INDW and AGRW.

- Canada and Switzerland 1830 to 1870, estimated CPI keeping the same wholesale to consumer prices proportion than in the period 1890-1910.

- Norway 1870 set up at 85 (1890 level), in light of the stability of prices.

*INDW = Real wage index in industry. INDWCURR/PRICES*100. This is, the index of nominal wages has been divided by the price index, and then multiplied by 100 in order to go back to the original scale of the variables.

*AGRW = Real wage index in agriculture. AGRWCURR/PRICES*100. This is, the index of nominal wages has been divided by the price index, and then multiplied by 100 in order to go back to the original scale of the variables.

*WEIGHTEDW = Weighted average of real industrial wages (INDW) and real agricultural wages (AGRW), weights according to the proportion of workers in each sector (INDLAB and AGRICLAB, respectively). This variable complements the variable WAGES for all those countries where Williamson's wage series are not available.

WAGES = real wages (Williamson, 1999, 2000a, and 2000b). When these are not available, they have been calculated from Mitchell's current wages by sector and corrected by the consumer price index (Mitchell, 2003a, 2003b, 2003c). If wages for more than one sector available, these are a weighted average (weights according to the proportion of workers in each sector).

Notes: - Egypt 1850 corresponds to 1858.

- Argentina 1850 corresponds to 1864.

- Wages for Brazil 1830 and 1850 are for Rio de Janeiro; from then onwards they are a simple average of Rio de Janeiro and Pernambuco.

- Germany 1870 has been extrapolated.

- New Zealand 1870 and 1890 have been set to 1910 level, due to the lack of data. Though there is no warranty that the real wage was constant, the continued stability of prices during this period justifies the decision.

- China 1870 and 1890 have been estimated keeping the average proportion of growth of the rest of countries in the sample.

- For Netherlands, Russia, Spain, and Switzerland we cannot make an informed decision, so an arbitrary real wage index of 100 has been assigned for 1900 (following Williamson), and then the average growth rate of real wages in the sample has been applied.

*POPGROUP = Classification for total population. Numerical codification: From 100 to 0 corresponding to A to G. The revision incorporates the previously unused table for 1914. Also available for 1830 (table A25 in Morris and Adelman, 1988:386). POP from more recent sources used instead.

POP = Total population in thousands (from Maddison, 2003), updated with his revised version dated August 2007 (http://www.ggdc.net/maddison/). Figures for non-available years have been interpolated at constant growth rate.

*POPGRGRP = Classification for rate of population growth in the last 20 years. Numerical codification: From 100 to 0 corresponding to A to F. POPGR used instead.

POPGR = Cumulative population growth in the last 20 years (%). It has been calculated from total population (POP).

*POPGRN = Total population growth, in levels (thousands). Calculated from total population (POP). This variable is necessary to calculate the rate of migration over total population growth (MITOPOPGR).

*IMMIGRP = Classification for net immigration. Numerical codification: From 100 to 0 corresponding to A to F. IMMI used instead.

IMMI = Net migration, in thousands (immigration with positive sign and emigration with negative sign). Source: Own calculations from Morris and Adelman (1988: Table A32). For cases where no statistics are available, we rely on the IMMIGRP classification, combined with the figures available for other years, to provide an estimate.

MITOPOPGR = Rate of net migration over total population growth (%). This variable has been calculated from net migration (IMMI) and total population growth, in levels (POPGRN).

LIT = Classification of extent of adult literacy. Re-codification: Enhanced precision scale from 0 to 100 corresponding to A to J in the original. Notice that the order of the numerical classification has been reversed in order to transform the illiteracy variable into literacy. This is intuitively convenient since we assign higher values to better outcomes in this way.

PRIMEDGR = Classification for rate of spread of primary education in the past 20 years. Numerical codification: From 100 to 0 corresponding to A to E.

*NPUPILS = Number of pupils in schools, in thousands (from Mitchell, 2003a, 2003b, and 2003c). This variable is necessary to calculate QEDU.

Note: - Sweden 1850 and Argentina 1870 correspond to own calculations, trusting the steady tendencies.

*NTEACHERS = Number of teachers in schools, in thousands (from Mitchell, 2003a, 2003b, and 2003c). This variable is necessary to calculate QEDU.

*NSECOND = Number of students in secondary education, in thousands (from Mitchell, 2003a, 2003b, and 2003c).

Note: For countries where there is no records of secondary schooling *and* primary school numbers are very low, we assume negligible secondary schooling.

*QEDUSEC = Secondary education students per thousand pupils in primary education. (NSECOND/NPUPILS)*1000. This variable has been devised to address the quality of education issue for Africa, Asia and Oceania, where there are no records of number of teachers. The way to address this problem has been to monitor the rate of students enrolled in secondary school to the number of pupils in primary school. The higher the proportion of secondary school with respect to primary one, the higher the quality of education.

Note: Changes in this variable do not coincide with changes in the number of teachers per thousand pupils (QEDU) (correlation between the two is to be around 0); so it has to be regarded as an alternative.

*QEDU = Teachers per thousand pupils. (NTEACHERS/NPUPILS)*1000, this is, the ratio of teachers to pupils multiplied by a thousand.

Notes: - Closest year used when particular year not available, trying to match number of teachers and number of students corresponding to the same year.

- UK data correspond to England and Wales.

- For some countries, figures include also secondary schooling, due to the inability to separate primary from secondary school records. The distortion

introduced should be fairly small, since secondary schooling at that time still represented a small proportion of the total.

LANDTENU = Classification for predominant form of land tenure and holding. Numerical codification: From 100 to 0 corresponding to A to G.

LANDCONC = Classification for concentration of landholdings. Numerical codification: From 100 to 0 corresponding to A to G.

LANDADOP = Classification for favourableness of land system to adoption of improvements. Numerical codification: From 100 to 0 corresponding to A to I.

URBANI = Classification for extent of urbanisation. Numerical codification: From 100 to 0 corresponding to A to D.

ENTREP = Classification for favourableness of attitudes towards entrepreneurship. Numerical codification: From 100 to 0 corresponding to A to F.

*GOVT = Classification for extent of domestic economic role of government in the past 20 years. Numerical codification: From 100 to 0 corresponding to A to E. GOVREL used instead.

*GOVTOTAL = Total central government expenditure, in millions of the local currency (from Mitchell, 2003a, 2003b, and 2003c). This variable is used to calculate public expenditure relative to GDP (GOVREL).

Notes: - Russia 1830 using the exchange rate of 26.3 paper rubles for a silver ruble for that year according to Lindert (2006:Table 3).

- Sweden 1870 has been extrapolated trusting the very stable progression in the government spending.

GOVREL = Public expenditure as a percentage of GDP. Calculated as GOVTOTAL/CURRENTGDP*100.

Note: - For cases where there is not enough data, either on government expenditure or current GDP, the old Adelman and Morris score for domestic economic role of the government (GOVT) has been replaced by a government expenditure level corresponding to the percentile matching with the original Adelman and Morris score.

SOCIOPOL = Classification for socioeconomic character of national political leadership in the past 20 years. From 100 to 0 corresponding to A to D.

*SCOPE = scope of government action. Source: Polity II (no longer available in later versions of the polity project). Even though it has been compiled and recorded in the database, this variable has been excluded from the principal components analysis due to poor variability of the data. The original variable socio-economic

character of political leadership (SOCIOPOL), has been used instead for the quantitative analysis.

Notes: - The polity II database uses 3-digit country codes. The codes are available from Gurr (1997, Appendix A:52-68).

- Argentina 1830 corresponds to 1835.
- Australia 1870 and 1890 correspond to 1901, where recording starts.
- Belgium 1830 corresponds to 1831, where recording starts.
- British colonies or British settlements, Australia 1830 and 1850, Burma, India, Canada 1830 and 1850, and Egypt 1890 and 1910, have been assigned UK scores.
- Denmark 1910 corresponds to the score before and after the disruptive period, which was the same score (before and until 1900, and 1915 and after).
- France 1870 corresponds to 1869.
- New Zealand 1850 corresponds to 1857.
- Spain 1870 corresponds to the score before and after the 2-year disruption.
- Sweden 1910 corresponds to the score before and after the disruption.
- Switzerland 1870 corresponds to the score before and after the interruption of data.
- Germany before becoming a nation-state corresponds to the average of Baden, Bavaria, Saxony, and Wuerttemburg.
- Italy before becoming a nation-state corresponds to the average of Modena, Parma, Sardinia, Tuscany and Two Sicilies.

*REPRESEN = Classification for strength of national political institutions in the past 20 years. From 100 to 0 corresponding to A to D.

POLITY2 = Revised Polity variable score, ranging from -10 to 10. Source: Polity IV, variable under the same name.

Notes: - Australia was given score 10 throughout the entire twentieth century in the Polity IV database. Therefore, the best guess for unavailable previous years is score 10.

- Burma was administered as a province of India until 1938, and gained independence from the Commonwealth in 1948 (source: The CIA Factbook), when the Polity IV database states reporting as Myanmar. In the case of New Zealand (see below), we see that prolonged periods of occupation in the nineteenth century do not seem to be disruptive for the polity scores; so, as for India, the scores assigned are those corresponding to the metropolis. And,

indeed, British and Burma scores remain close for a long time after independence until 1961, when the first military regime started (*ibid.*).

- Canada data for years prior to 1867 have been assigned value 4, which corresponds to the 1867-1887 score.

- Polity IV reporting for Egypt starts in 1922, at the time of partial independence from the UK. At that time, Egyptian score for the polity variable is 4 while the British score has been 8 since the turn of century. It seems reasonable then to assign Egypt 1910 the 1922 score. Previous scores for Egypt have had to be arbitrary, but follow the same proportion to the British score as that of the time of partial independence. The Turkish (previous invader) score remain at its minimum value -10 throughout the nineteenth century. But during the Ottoman occupation of Egypt the Mamluck continued governing the country (source: The CIA World Factbook), so there are no reasons to believe that Egypt would have had such low polity scores. Proportions to the British score have been kept throughout.

- Germany 1830 and 1850 correspond to Prussia.

- India has been given the same scores as the metropolis. They still show a very close score since time of independence, when separate Indian records start in the Polity IV database.

- Italy 1830 and 1850 correspond to the invariant nineteenth century score starting at the time of unification, 1861.

- The same score, 10, is given in the Polity IV database for New Zealand from 1857 to 1875, with no records before the initial date of this period. A series of land wars between the first colonizers, the Maoris, and the newly arrived British colonizers took place between 1843 and 1872 (source: The CIA World Factbook). However, this fact seems not to have affected the polity score at all. Therefore, the same score is maintained for 1830 and 1850, corresponding to the years surrounding the arrival of the first British colonizers.

- Switzerland 1830 has been assigned a score of 10, in views of the stunning homogeneity of the series.

POLSTABI = Classification for extent of political stability in the past 20 years. Codification: From 100 to 0 corresponding to A to D.

FOREIGND = Classification for degree of foreign economic dependence in the past 20 years. Inverted numerical Re-codification: Enhanced precision scale From 0 to 100 corresponding to A to G. We assign a higher punctuation to economic

independence, unlike the original. This eases the conceptualisation of the variable, since economic dependence is taken to be negative.

COLSTAT = Classification for colonial status. Numerical Re-codification: Enhanced precision scale Inverted from 0 to 100 corresponding to A to F.

Finally, the market institutional development variables come from an earlier study by the same authors. They selected the following three composite indicators, which summarise the level of institutional development of markets. For a detailed explanation, visit Adelman and Morris (1978).

MKTDEV = Component scores for composite indicator of level of development of market institutions up to the given date.

MKTDEVGR = Component scores for composite indicator of rate of spread of market institutions in the last 20 years.

MKTDVGRL = Component scores for composite indicator of rate of spread of market institutions in the last 20 years (lagged); i. e., the observation corresponding to 1850 is capturing the spread of market institutions from 1830 to 1850.

TABLES

Table 2.1. **Value of Exports at Constant Prices (35 Countries), 1820–1913** (million 1990 dollars)

	1820	*1830* (my own calculations)	*1850* (my own calculations)	*1870*	*1890* (my own calculations)	*1910* (my own calculations)	*1913*
Austria	47	74,39	186,39	467	923,73	1827,16	2024
Belgium	92	147,28	377,47	967,43	2479,42	6354,51	7318
Denmark			152,00	314	648,65	1339,95	1494
Finland			144,62	310	664,50	1424,41	1597
France	487	722,99	1593,47	3512	6046,01	10408,39	11292
Germany			3021,48	6761	15128,74	33852,80	38200
Italy	339	448,94	787,37	1380,91	2421,88	4247,55	4621
Netherlands			1126,33	1727	2648,00	4060,17	4329
Norway			119,42	223	416,43	777,63	854
Sweden							2670
Switzerland	147	220,13	493,65	1107	2379,14	5113,20	5735
United Kingdom	1125	1813,25	4710,49	12237	21067,07	36268,81	39348
Total	n.a.		16080,94	30396	57454,17	108599,22	119482
Australia			178,74	455	1158,24	2948,41	3392
Canada			325,28	724	1611,44	3586,66	4044
United States	251	397,33	995,66	2495	6445,09	16648,96	19196
Total	n.a.		1462,23	3674	9231,29	23194,52	26632
Spain	137	197,36	409,58	850	1684,08	3336,63	3697
USSR				n.a.			6666
Argentina			80,55	222	611,81	1686,09	1963
Brazil			590,48	854	1235,13	1786,34	1888
Chile			84,89	166	324,62	634,81	702
Colombia			76,74	114	169,36	251,61	267
Mexico			83,85	242	698,42	2015,67	2363
Peru			145,50	202	280,47	389,36	409
Venezuela				n.a.			1374
Total				2126	4152,19	8109,45	8966
Bangladesh				-			-
Burma				-			-
China			838,39	1398	2331,14	3887,14	4197
India			2170,61	3466	5534,46	8837,34	9480
Japan	51						
Pakistan				-			-
Philippines				55	95,47	165,71	180
South Korea				0		0	171
Taiwan				-			70
Thailand				88	196,51	438,80	495
Total				5230	9114,92	15885,62	17266

Source: Maddison (2006:Table F-2) and own calculations

Table 2.2. **Correlations between Education Variables**

		Literacy	Primedgr	N pupils	N teachers	N second	Qsec edu	Q edu
Literacy	Pearson Correlation	1	.668(**)	.176	.150	-.350	-.281	.146
	Sig. (2-tailed)		.000	.225	.446	.058	.102	.459
	N	69	69	49	28	30	35	28
Primary	Pearson Correlation	.668(**)	1	.154	.000	-.109	-.329	-.268
Education	Sig. (2-tailed)	.000		.290	.999	.568	.053	.168
Growth	N	69	69	49	28	30	35	28
Number of	Pearson Correlation	.176	.154	1	.953(**)	.741(**)	.009	-.029
pupils	Sig. (2-tailed)	.225	.290		.000	.000	.951	.851
	N	49	49	73	43	45	54	43
Number of	Pearson Correlation	.150	.000	.953(**)	1	.667(**)	.040	.113
teachers	Sig. (2-tailed)	.446	.999	.000		.000	.820	.469
	N	28	28	43	43	27	35	43
Number in	Pearson Correlation	-.350	-.109	.741(**)	.667(**)	1	.454(**)	-.046
secondary	Sig. (2-tailed)	.058	.568	.000	.000		.002	.820
	N	30	30	45	27	47	45	27
Quality sec	Pearson Correlation	-.281	-.329	.009	.040	.454(**)	1	-.087
education	Sig. (2-tailed)	.102	.053	.951	.820	.002		.621
	N	35	35	54	35	45	54	35
Quality	Pearson Correlation	.146	-.268	-.029	.113	-.046	-.087	1
education	Sig. (2-tailed)	.459	.168	.851	.469	.820	.621	
	N	28	28	43	43	27	35	43

** Correlation is significant at the 0.01 level (2-tailed).

Table 2.3. **Percentiles of Total Central Government Expenditure Relative to GDP**

countryear	Govrel	Cumulative percent
Switzerl 1850	.37	1.5
India 1830	.83	3.0
Switzerl 1870	.94	4.5
Inferred value (China 1870 China 1890 China 1910)	**1.06**	**5**
Switzerl 1890	1.31	6.0
India 1850	1.31	7.5
Belgium 1830	1.43	9.0

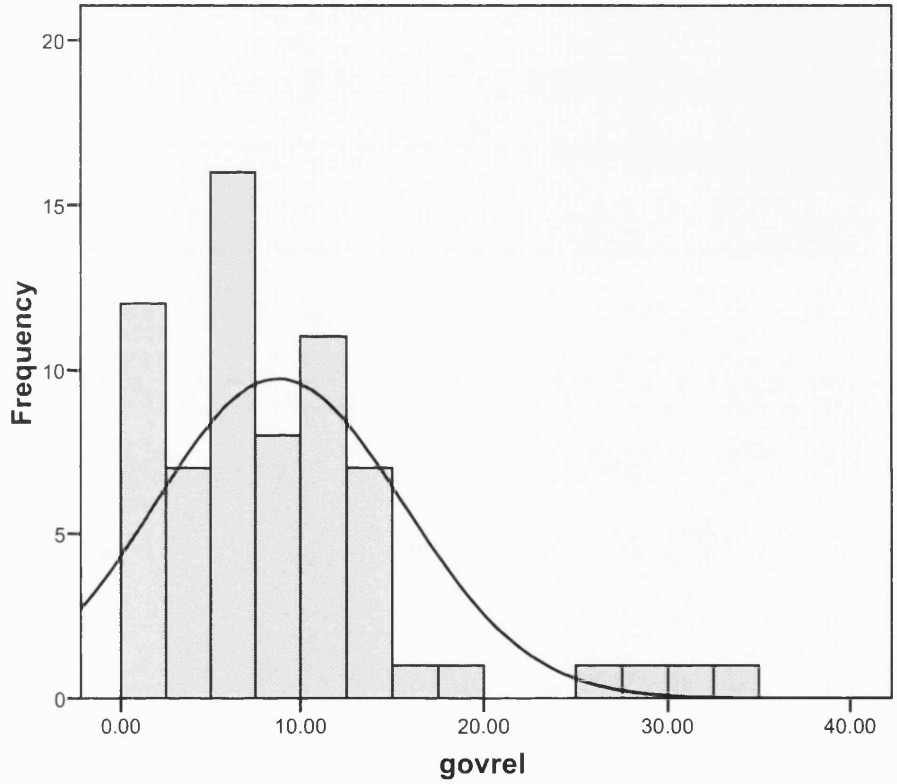
Inferred value (Burma 1870 Burma 1890 New Zeal 1870 Norway 1870 Norway 1890)	**1.52**	**10**
US 1850	1.55	10.4
US 1830	1.61	11.9
US 1910	1.97	13.4
Switzerl 1910	2.30	14.9
India 1870	2.42	16.4
US 1890	2.43	17.9
Japan 1870	2.60	19.4
India 1890	3.24	20.9
Canada 1870	3.93	22.4
Belgium 1850	4.03	23.9
India 1910	4.19	25.4
US 1870	4.22	26.9
Canada 1890	4.91	28.4
Belgium 1870	5.04	29.9
Inferred value (Argentina 1870 Argentina 1890 Argentina 1910 Burma 1910 Egypt 1870 Egypt 1890 New Zeal 1890 Russia 1870)	**5.09**	**30**
Canada 1910	5.71	31.3
Germany 1890	5.72	32.8
Spain 1850	6.02	34.3
UK 1870	6.30	35.8
Denmark 1890	6.42	37.3
Germany 1910	6.60	38.8
Denmark 1870	6.73	40.3
UK 1890	6.77	41.8
Sweden 1890	6.80	43.3
Sweden 1870	6.83	44.8
Australia 1910	6.90	46.3

Denmark 1910	6.92	47.8
Brazil 1830	6.94	49.3
Inferred value (Norway 1910 Russia 1890)	**7.05**	**50**
Sweden 1910	7.16	50.7
Australia 1850	7.31	52.2
Netherlan 1850	7.53	53.7
Belgium 1890	7.62	55.2
Japan 1890	7.77	56.7
UK 1910	8.19	58.2
Netherlan 1910	8.26	59.7
Netherlan 1870	8.90	61.2
Spain 1910	8.98	62.7
France 1830	9.59	64.2
Australia 1870	10.16	65.7
Spain 1890	10.29	67.2
UK 1850	10.30	68.7
Inferred value (Egypt 1910 Russia 1910)	**10.32**	**70**
France 1850	10.32	70.1
France 1910	10.56	71.6
Brazil 1850	10.74	73.1
Germany 1870	10.93	74.6
Spain 1870	11.27	76.1
France 1890	11.36	77.6
UK 1830	11.82	79.1
Netherlan 1890	12.39	80.6
Belgium 1910	12.75	82.1
Brazil 1910	13.00	83.6
Inferred value (New Zeal 1910)	**13.22**	**85**
France 1870	13.24	85.1
Italy 1870	13.28	86.6
Italy 1910	13.75	88.1
Italy 1890	14.57	89.6
Italy 1850	14.71	91.0
Denmark 1850	16.19	92.5
Australia 1890	17.57	94.0

Brazil 1890	25.70	95.5
Brazil 1870	28.98	97.0
Japan 1910	31.16	98.5
Australia 1830	34.59	100.0

Sources: Govrel is total central government expenditure as a percentage to GDP, derived from Mitchell (2003a:886-908, 2003b:664-78, and 2003c:816-25). Values in bold are inferred from the percentiles, corresponding to domestic economic role of the government (Morris and Adelman: 1988: Table A42).

Histogram of Total Government Expenditure Relative to GDP



REFERENCES

Adelman, Irma, and Cynthia Taft Morris, 1978, 'Paterns of Market Expansion in the Nineteenth Century: A Quantitative Study', in George Dalton, ed., *Research in Economic Anthropology*, JAI Press (Greenwich, Connecticut):159-176.

Alesina, Alberto, and Roberto Perotti, 1996, 'Income Distribution, Political Instability, and Investment', *European Economic Review*, 40:1203-1228.

Anderson, Perry, 1974, Lineages of the Absolutist State, New Left Books (London).

Ashworth, William, 1962, *A Short History of the International Economy since 1850*, 2nd edition, Longmans, Green (London).

Bairoch, Paul, 1968, 'La Population Active et sa Structure', in Tilo Deldycke, H. Gelders, and Jean Marie Limbor, eds. , Series: *International Historical Statistics*, vol I, Centre d'Économie Politique de l'Université Libre de Bruxelles (Brussels).

—— , 1969, *Diagnostic de l'Évolution Économique du Tiers-Monde 1900-1968*, Gathier-Villars (Paris).

—— , 1976, 'Europe's Gross National Product: 1800-1975', *Journal of European Economic History* 5 (Fall):273-340.

Bairoch, Paul and Maurice Levy-Leboyer, ed., *Disparities in Economic Development since the Industrial Revolution,* The MacMillan Press Ltd., (London and Basingstoke, 1981).

Banks, Arthur S., 1971, *Cross-Polity Time-Series Data*, The MIT Press (Cambridge, MA, and London, UK).

—— , 2005, *Cross-National Time-Series Data Archive*, http://www.databanks.sitehosting.net/ (last accessed 05/12/2007).

Banks, Arthur S., and Robert B. Textor, 1963, *A Cross-Polity Survey*, The MIT Press (Cambridge, MA).

Barro, Robert J., 1991, 'Economic Growth in a Cross Section of Countries', *Quarterly Journal of Economics*, 106 (2): 407-443.

Barro, Robert J. and Jong-Wha Lee, 'International Measures of Schooling Years and Schooling Quality', *American Economic Review, Papers and Proceedings* 86 (2), 1996, pp. 218-23. (Updated dataset at: http://www.worldbank.org/research/growth/ddbarle2.htm)

Boserup, Mogens, 1963, 'Agrarian Structure and Take-off' in Walt Whitman Rostow, ed., *The Economics of Take-off into Sustained Growth*, Proceedings of the International Economic Association conference, Macmillan (London):201-224.

Cairncross, Alexander Kirkland, 1962, *Factors in Economic Development*, Allen & Unwin (London).

Center for International Development and Conflict Management, *Polity IV Project* http://www.cidcm.umd.edu/polity/ (last accessed 27/11/2007).

Central Intelligence Agency, 2007, *The World Factbook*, https://www.cia.gov/library/publications/the-world-factbook/ (last accessed 29/11/2007).

Cipolla, Carlo M., 1974, *The Economic History of World Population*, 6th edition, Penguin Books (Baltimore).

Clapham, John Harold, 1936, *The Economic Development of France and Germany, 1815-1914*, 4th editon, Cambridge University Press (Cambridge).

Clark, Colin, 1957, *The Conditions of Economic Progress*, 3rd edition, McMillan (London).

Clough, Shepard Bancroft, and Charles Woolsey Cole, 1946, *Economic History of Europe*, Revised edition, Heath (Boston).

Cohen, Daniel and Marcelo Soto, 'Growth and Human Capital: Good Data, Good Results', CEPR discussion paper 3025, (2001).

Crafts, Nicholas and Anthony J. Venables, 'Globalization in History: A Geographical Perspective', CEPR discussion paper 3079, (2001).

De la Fuente, Angel and Rafael Doménech, 'Human Capital in Growth Regression: How Much Difference Does Quality Data Make', CEPR discussion paper 2466, (2000).

Easterly, William, and Ross Levine, 1997, 'Africa's Growth Tragedy: Policies and Ethnic Divisions', *Quarterly Journal of Economics*, 112 (4): 1203-50.

Eichengreen, Barry, 1986, 'What Have We Learned from Historical Comparisons of Income and Productivity?', in Patrick O'Brien, ed., *International Productivity Comparisons and Problems of Measurement, 1750–1939,* 9th International Economic History Congress, Session B6, Bern:26–35.

Engerman, Stanley L., and Kenneth L. Sokoloff, 'Factor Endowments, Institutions, and Differential Paths of Growth Among New World Economies', in S. Haber, ed., *How Latin America Fell Behind: Essays on the Economic Histories of Brazil and Mexico, 1800-1914* , (Stanford University Press, 1997), Chapter 10, pp. 260-291.

Ferenczi, Imre, and Walter F. Willcox, eds., 1929, *International Migrations*, vol. I, *Statistics*, Publications of the NBER 14 and 18 (New York).

Flora, Peter, *State, Economy, and Society in Western Europe 1815-1975: A Datahandbook in Two Volumes* , Vol I. 'The Growth of Mass Democracies and Welfare State', Campus Verklklag (Frankfurt, 1983).

—— , *State, Economy, and Society in Western Europe 1815-1975: A Datahandbook in Two Volumes*, Vol II. 'The Growth of Industrial Societies and Capitalist Economies', Campus Verlag (Frankfurt, 1987).

Furnival, John S., 1931, *An Introduction to the Political Economy of Burma*, Burma Book Club (Rangoon).

Gerschenkron, Alexander, 1962, *Economic Backwardness in Historical Perspective: A Book of Essays*, Harvard University Press (Cambridge, MA).

Girard, L., 1965, 'Transport', in H. J. Habakkuk and M. M. Postan, eds., *The Cambridge Economic History of Europe*, Vol. VI, *The Industrial Revolution and After: Incomes, Population, and Technological Change*, Cambridge University Press (Cambridge).

Gurr, Ted Robert, 1997, 'Polity II: Political Structures and Regime Change, 1800-1986', Inter-university Consortium for Political and Social Research (ICPSR) 9263.

Henisz, Witold Jerzy, 2000, 'The Institutional Environment for Economic Growth' *Economics and Politics*, 12(1):1-31.

Henisz, Witold Jerzy, and Bennet A. Zelner, 'Measures of Political Risk', in *The Political Constraints Index (POLCON) Dataset*, http://www-anagement.wharton.upenn.edu/henisz/POLCON/ContactInfo.html (last accessed 05/12/2007).

Heston, Alan, Robert Summers and Bettina Aten, Penn World Table Version 6.1, Center for International Comparisons at the University of Pennsylvania (CICUP), October 2002.

Isenman, P., 1980, 'Inter-Country Comparison of 'Real' (PPP) Incomes: Revised Estimates and Unresolved Questions', *World Development,* 8:61–72.

Kaufmann, Daniel, Aart Kraay and Massimo Mastruzzi, 2007, July, 'Governance Matters VI: Governance Indicators for 1996-2006', World Bank Policy Research Working Paper 4280.

Knack, Stephen, and Philip Keefer, 1997, 'Does Social Capital Have an Economic Payoff? A Cross-Country Investigation', *The Quarterly Journal of Economics*, CXII :1251-1288.

Krueger, Alan B. and Mikael Lindahl, 'Education for Growth: Why and For Whom', NBER working paper 7591 (2000).

Landefeld, J. Steven, and Robert P. Parker, 1997, 'BEA's Chain Indexes, Time Series, and Measures of Long Term Economic Growth', *Survey of Current Business*, 77 (5, May), html version.

Lee, William Robert, 1977, Population Growth, Economic Development, and Social Change in Bavaria, 1750-1859, Arno Press (New York).

Lindert, Peter, 2006, 'Silver and gold content of the Russian ruble, 1535-1913', Global Price and Income History Group http://gpih.ucdavis.edu/files/Russia_Ag_content_ruble_1535-1913.xls (last accessed 4/11/2007).

Maddison, Angus, 1979, 'Per Capita Output in the Long Run', *Kyklos* 32,1-2:412-29.

——, 1982, *Phases of Capitalist Development*, Oxford University Press (New York).

——, 1995, *Monitoring the World Economy: 1820-1992*, OECD Development Centre (Paris).

——, 1998, *Chinese Economic Performance in the Long Run,* OECD Development Centre (Paris).

——, 2001, *The World Economy: A Milenial Perspective*, OECD Development Centre (Paris).

——, 2003, *The World Economy: Historical Statistics*, OECD Development Centre (Paris).

——, 2006, *The World Economy*, (includes Volume 1: A Millennial Perspective and Volume 2: Historical Statistics), OECD Development Centre (Paris).

——, 2007, August, *World Population, GDP and Per Capita GDP, 1-2003 AD*, Angus Maddison Homepage http://www.ggdc.net/maddison/ (last accessed 27/09/2007).

Marshall, Monty G. and Keith Jaggers, 2005, 'Polity IV Project: Political Regime Characteristics and Transitions, 1800-2004. Data User's Manual', Center for Global Policy (Arlington, VA).

Milward, Alan S., and S. B. Saul, 1979, The Economic Development of Continental Europe, 1780-1870, 2nd ed, George Allen & Unwin, (London).

Mitchell, Brian Redman, 1973, 'Statistical Appendix', in Carlo M. Cipolla, ed., *The Fontana Economic History of Europe*, Vol IV, *The Emergence of Industrial Societies*, Collins/Fontana Books (London):738-820.

——, 1976, *European Historical Statistics, 1750-1970*, Columbia University Press (New York).

——, 1998a, *International Historical Statistics: Europe 1750-1993*, Fourth edition, Stockton Press and MacMillan Reference Ltd, (New York and Basingstoke

respectively).

—— , 1998b, *International Historical Statistics: The Americas 1750-1993*, Fourth edition, Stockton Press and MacMillan Reference Ltd, (New York and Basingstoke respectively).

—— , 2003a, *International Historical Statistics: Africa, Asia, and Oceania 1750-2000*, Fourth edition, Palgrave MacMillan, (Basingstoke).

—— , 2003b, *International Historical Statistics: The Americas, 1750-2000*, Fifth edition, Palgrave McMillan, (Basingstoke).

—— , 2003c, *International Historical Statistics: Europe, 1750-2000*, Fifth edition, Palgrave McMillan, (Basingstoke).

Moore, Barrington, Jr., 1966, *Social Origins of Dictatorship and Democracy: Lord and Peasant in the Making of the Modern World*, Beacon Press (Boston).

Moore, William H., Polity Data Page: http://garnet.acns.fsu.edu/~whmoore/polity/polity.html (accessed 04/06/2003).

Morris, Cynthia Taft and Irma Adelman, *Comparative Patterns of Economic Development, 180-1914*, The John Hopkins University Press (Baltimore and London, 1988).

Mulhall, Michael G., 1899, *The Dictionary of Statistics*, 4th edition, Routledge & Sons (London).

Mukerjee, Tapan, 1972, 'Theory of Economic Drain: impact of British Rule on the Indian Economy, 1840-1900' in *Economic Imperialism: A Book of Readings.*

Pollard, Sidney, 1988, Review on *Comparative Patterns of Economic Development, 1850-1914*, by Cynthia Taft Morris and Irma Adelman, 1988, *Economic History Review*, 41 (4, November):668-670.

Prados de la Escosura, Leandro, 'International Comparisons of Real Product, 1820–1990: An Alternative Data Set', *Explorations in Economic History* 37 (2000), pp. 1–41.

Putnam, Robert D., *Making Democracy Work: Civic Traditions in Modern Italy*, (Princeton University Press, 1993).

—— , *Bowling Alone: The Collapse and Revival of American Community*, Simon & Schuster, (New York, 2000).

Roe, Mark J., and Jordan I. Siegel, 2007, *Political Instability and Financial Development*, Social Science Research Network (SSRN) http://ssrn.com/abstract=963214 (last accessed 05/12/2007).

Schimidt-Hebbel, Klaus, 1996, 'Fiscal Adjustment and Growth: In and Out of Africa',

Journal of African Economies, 5 (3): 7-59.

Summers, Robert and Alan Heston, 1988, 'A new set of international comparisons of real product and price levels estimates for 130 countries, 1950-1985', *Review of Income and Wealth*, 34(1), March:1-25.

Temple, Jonathan, and Paul A. Johnson, 'Social Capability and Economic Growth', *The Quarterly Journal of Economics* 113 (3), 1998, pp. 965-990.

The Statesman's Yearbook: The Politics, Cultures and Economies of the World. Yearly periodical since 1864, McMillan (London).

Thornthwaite, Charles Warren, 1933, (July) 'The Climates of the Earth', *Geographical Review* 23 (3):433-440.

United Nations Population Division, 2001, *World Population Prospects: The 2000 Revision*, Department of Economic and Social Affairs (New York).

Warren, Bill, 1980, *Imperialism: Pioneer of Capitalism*, New Left Books (London).

Watson, Peter D., Centre for Conflict and Cooperation data web page: http://www.watson.rochester.edu/resources/data.html (accessed 04/06/2003).

Weber, Adna Ferrin, 1967, *The Growth of Cities in the Nineteenth Century: A Study in Statistics*, Cornell University Press (Ithaca, NY).

Williamson, Jeffrey G., 1999, 'Real Wages, Inequality, and Globalization in Latin America Before 1940', *Revista de Historia Economica* 17, special number: 101-142.

——, 2000a, 'Real Wages and Factor Prices Around the Mediterranean 1500-1940', in S. Pamuk and J.G. Williamson, ed., *The Mediterranean Response to Globalization Before 1950* (London: Routledge), pp. 45-75.

——, 2000b, 'Globalization, Factor Prices and Living Standards in Asia Before 1940', in A.J.H. Latham and H. Kawakatsu, ed., *Asia Pacific Dynamism 1500-2000* (London: Routledge), pp. 13-45.

World Bank Group, The, *The Economics of Civil War, Crime and Violence: Related Sites*: http://econ.worldbank.org/programs/conflict/topic/12377/ (accessed 04/06/2003).

Chapter 3

A SOCIAL DEVELOPMENT INDEX FOR THE NINETEENTH CENTURY

Is Social Capital Persistent? Comparative Measurement in the Nineteenth and Twentieth Centuries

ABSTRACT

Recently, there has been a growing interest in social capital and in the difficulties related to its measurement. This paper proposes to measure social capital by means of principal components analysis and presents the first available international social capital estimates for the nineteenth century. The analysis is based on a nineteenth-century international database containing a wide range of socio-economic variables. Social capital indicators are constructed for the years 1870 and 1890. Interestingly enough, these indicators are comparable to mid-twentieth century social indicators. This facilitates the study of the evolution of social capital between the nineteenth and twentieth centuries. In the very long run, one can find a significant decline in the relative position of the European countries and the United States.

JEL classification codes: Z13, N01, O11

3. I. INTRODUCTION

A recent stream in the economics literature argues that history affects economic performance through institutions. Perhaps the most well-known paper is that by Acemoglu, Jonhson, and Robinson (2001), where they argue that different colonial origins led to different sets of institutions, and the latter gave rise to different economic outcomes. But, what happens when we look at colonisers instead of colonised countries? Is there nothing to say? Guido Tabellini (2007) conducts an empirical investigation on differential economic performance in the European regions, and links this fact to deep cultural roots, based on trust, respect and confidence built up centuries ago. Both these branches represent two separated parts of the world. On the one hand, the colonisers acquiring their culture through centuries of social interaction; on the other hand, the colonisers, inheriting the institutional bequest transferred by the colonisers. Now, is it possible to conduct a historical investigation with a sample of countries representing all regions of the world? And, more importantly, can one reach a generalised conclusion of whether culture and institutions are deterministic or not so when it comes to economic performance? This chapter proposes an international index for social capital in the nineteenth century –very close to Tabellini's definition of culture–, which comprises a varied sample of countries around the world, including both colonised and colonisers. The next chapter enquires about its potential economic influence.

Recapitulating, the first chapter provided a definition of social capital and embedded it in the economics literature, the way it has been studied until now. In a phrase, social capital is the quality of the civil society, or the civic values. It should interest economists because it is indeed a form of capital[1]. Social capital has been shown to have had an effect in economic performance for the recent decades, but there are some indications that it could also have been relevant in the distant past. Chapter 1 gives some historical examples that set the ground to believe that it is promising to give social capital a long run approach. Actually, in his investigation for the European regions, Tabellini (2007) concludes that economic outcomes have their root in historically fundamented attitudes towards trust, respect for others and self-determination. Nevertheless, it is hard to device quantitative tests. The second chapter presents a completely renovated version of an old socio-economic database, which is going to turn

[1] See chapter 1- introduction

useful in the estimation of social development levels towards the end of the nineteenth century. This chapter motivates and explains the statistical methodology used for that purpose, and presents the new series. Also, it compares the new (referring to the past) to the existing (more recent) indicators and assesses the viability and performance of the different measurement alternatives.

More challenging questions are: How does social capital evolve over time and across countries? Is it path dependent? Adding some time dimension to the study of social capital looks promising. Section 3.II compares three different proposed measures of social capital for the second half of the twentieth century, and looks at the relationships between them. Section 3.III turns the attention into historical data and presents a new social capital index for the late nineteenth century. In Section 3.IV, we are able to look at the inter-temporal evolution of social capital thanks to the newly created index, together with the more recent measurement alternatives. Finally, section 3.V succinctly concludes.

3. II. COMPARISON OF CONTEMPORARY ALTERNATIVES

Of all the possible ways to try to measure social capital, several empirical proposals are available, but yet none enjoys general acceptance. Therefore, it is convenient to start with a comparison amongst the most popular measuring alternatives. The comparison might turn into an interesting exercise, since it will make us win some perspective on the alternatives plus reveal some of the insights that remain unnoticed until the present moment. This section compares three different twentieth century measurement alternatives, based on the pre-existing measurement attempts. These three are TRUST, CIVIC, and SOCDEV, standing for level of trust in a society, civic engagement, and social development respectively.

Both TRUST and CIVIC have been originally extracted from the World Value Surveys, which periodically runs over a whole range of countries over the world. General trust in people (TRUST) is the percentage of respondents who answered 'yes' to the following question: 'Generally speaking, would you say that most people can be trusted or that you can't be too careful in dealing with people?' I amplified TRUST to TRUSTAM by adding extra country data from both the most recent and past rounds of the World Value Surveys. Civic engagement (CIVIC) is the percentage of civic activities in which and

average individual participates. The activities included are: social-welfare services for elderly and deprived; education, art, and cultural activities; local community affairs; conservation, environment, ecology; and voluntary associations for health (La Porta et al., 1997). Finally, social development (SOCDEV) was taken from Adelman and Morris (1967). The index is an extraction of factor scores from a principal components analysis including 41 socio-economic variables from 74 countries around the world, for the period 1957-62. Temple and Johnson (1998) used this index before in order to test the economic significance of social arrangements.

With respect to the timing of the measures, the first two correspond to the late twentieth century and have been taken from Knack and Keefer (1997), (hence KK appears in some instances attached to the variable name). Overall, the General Value Survey rounds for 1980, 1990-1 and 1995-6 are incorporated, using always the most recent observation available[2]. As an average, we can say the two variables are aimed at monitoring social capabilities at the end of the twentieth century. SOCDEV corresponds to the early 1960's, –this is, twenty to thirty-five years earlier–, so we will need to keep this in mind.

Table 3.1[3] is a compendium of the data availability for three variables. We can observe that data availability is limited, so the comparisons amongst variables are forced to be restricted to a smaller sample of countries.

These are three different ways to measure social capabilities that have been proposed in the literature. They are conceptually different from each other and may or may not be related. The correlation matrix shows that their relationship, if any, is not always linear (table 3.2). Later we will find non-linear relationships between them.

In order to investigate deeper the relationships between the three variables, we make use of graphical representations. Only 9 complete cases are available for all three variables. This is due to the fact that the variables come from different sources and were not thought to match and be studied together. A scatter plot can help us to position these cases in the three-dimensional space. Figure 3.1 is a joint three-dimensional graphical representation of the three proposed indicators. SOCDEV stands for the Adelman and Morris social development index and is the variable positioned in one of the axis.

[2] The last round of Surveys dating 2002 is not included because it is beyond the time framework.
[3] Because of the large amount of calculations involved in this chapter, all tables are compiled at the end.

CIVICKK stands for civic engagement as reported by Knack and Keefer (1997) and has been placed in the second axis. Finally, TRUSTAM stands for trust amplified as described above and is the variable positioned in the third axis. Every dot in the three dimensional space reveals the position of the indicated country with respect to the three indicators. For best visualisation on the three-dimensional space, we include two scatter plots representing two different perspectives on the same data; –one with spikes to the floor and a second one with centroid perspective (spikes to the centre of the data) –. The three-dimensional graphical representation offers an overall picture of the data, which reveals an elliptic shape.

Since we are especially interested in historical considerations, the focus of the paper is driven by past-present contrasts. Nonetheless, comparing both contemporary measures for social capital, TRUST and CIVIC, is not of least interest, since they stand for different concepts. This is done by overlapping throughout the paper two plots in one. In the figures 3.2 and 3.3 (within twentieth century comparisons), the light coloured dots and lines depict the pair CIVIC *versus* SOCDEV, while the dark dots and lines represent TRUST *versus* SOCDEV. In short, the scatter plots below should be read in the following way: Every graph is composed of two overlapping bi-dimensional scatter plots, with the historical index in the horizontal axis and the contemporary index in the vertical axis. SOCDEV is common for both overlapping plots and is always positioned in the horizontal axis. It represents the historical measure of social capital. CIVIC and TRUST are always placed in the vertical axis, representing the contemporary measure of social capital. In this way, we can read all the graphs as a historical evolution of social capital, by looking at where countries were positioned in the 1960's (horizontal axis) and where they were positioned in the 1990's (vertical axis).[4]

Figures 3.2 and 3.3 represent the historical evolution of social capital in the second half of the twentieth century. SOCDEV corresponds to the years around 1960, while TRUST and CIVIC capture roughly the last two decades of the twentieth century. The two figures are based on the same data, and differ only on visual aids. The first one draws spikes from every country to the mean of the contemporary variable. Both variables CIVIC and TRUST have been standardised and thus vary within the same range. It is particularly interesting to observe where the mean of these two falls. We can observe

[4] STD at the end of the name of the variable means that the variable has been standardised. The variables which do not contain STD at the end of their name were already constructed in a way which allows for comparison.

that the mean of CIVIC is higher than the mean of TRUST. This fact can be due to the formulation of the questionnaire. But we should recall and keep in mind that they do not measure exactly the same concept: One is an index of voluntary participation and the other a percentage of 'yes' or 'no' answers regarding the general level of trust in a country. So there should be room for disparity. Nevertheless, it is still interesting to wonder whether there is a reason beyond formulation of the questionnaire and conceptual disparity behind the differing means. Later in this paper we argue that this is the case.

In figure 3.3 a line has been fitted to the points using a non-parametric technique called 'lowess' (locally weighted linear regression). This method fits the maximum number of points with the minimum number of iterations. Fifty percent of the points have been fitted with only three iterations. This type of graph is very appealing because it reveals the outliers. For the sake of historical findings, the engagingness of this exercise lays more on unmasking the outliers than on the fitted points that stand on the average. The impossibility of the fit line to match all dots points at the exceptional evolutions (both for the good and for the bad).

Striking results are those of India and Japan. They reveal themselves as outstanding performers in social improvement, which is historically consistent with their growth experiences. We can also detect failure stories by looking at the extremely poor contemporary scores compared to the mid-century scores for some Latin American countries like Mexico or Venezuela. Indeed, from figures 3.1 to 3.3 we can observe how some countries strikingly detach from the average, defeating the path dependence argument postulated by North. Having said this, the path dependence hypothesis is not refuted but modified. This is saying that socially well-endowed countries do actually leap over the development gap.

Should we have time series information about social evolution, we would be able to determine the timing of the social change: before, during, or after economic growth. As discussed in previous sections, Putnam argues that social change happens up to 70 years ahead of subsequent economic growth. Therefore, we need to go back further into history of social development to be able to contrast this observation. This is done in the next section.

3. III. CONSTRUCTING A SOCIAL CAPITAL INDEX FOR THE NINETEENTH CENTURY

Constructing a social capital measure for the distant past presents several challenges. The first challenge is the limitations of data availability. Once we turn into the distant past (more than a few decades ago) no surveys can be conducted and one has to rely on data already collected for other purposes. The second challenge, which is going to be addressed now, is to find a quantitative methodology that is flexible enough given the data limitations, but still conveys informative results. For both challenges, one needs to be truly imaginative and make the most out of the resources. Finally, how are the resulting data going to be compared to more recent data? The third challenge is to construct a measure that, at the same time, can be compared –even if imperfectly– to some existing indicator.

3. III. 1. SOURCES: NEW ADELMAN AND MORRIS DATABASE

Contemporary indicators of social capital based on the World Value Surveys are informative. Yet we need a wider time span in order to bring historical perspective into the analysis of social capital. Having pre-First World War social capital estimations would provide useful historical insights in order to study its evolution and test its persistence.

It is possible to find historical data to fill in the blanks on existing work and give a time dimension to the social capital analysis. At this respect, Adelman and Morris (1988) provide an extensive socio-economic database for the period 1850 to 1914.

The comprehensive nineteenth century series provided by Adelman and Morris are the starting point for our database. The extensive data appendix accompanying their 1988 book is a summary of the work on their data over more than 20 years. It contains cross-sectional data for 23 countries scattered over the globe and referring to 35 summary variables ranging from attitudes to change to political perceptions[5]. The latter depict the socio-economic structure of every country in the sample between 1850 and 1914, being this divided into 3 sub-periods: 1850-1870, 1870-1890, and 1890-1914. Cross-sectional

[5] A list of countries can be found in appendix 3.A; a detailed list of variables can be found in appendix 3.B.

data are supplied for every sub-period. The variables in levels and proportions refer to the initial level of each period, while those capturing change or characteristics refer to the whole of the preceding 20 year period.

The Adelman and Morris database has unique characteristics of which an economist looking for social influences in historical perspective can certainly take advantage of. These are: The database describes the situation of the economy in the late nineteenth century in conjunction with a detailed picture of the institutional framework, and some interesting social attitudes and customs in different countries. This highly valuable database has been explored under its possibilities.

This research reconstructs a similar database to that which Adelman and Morris built for the period 1850 to 1914, and then uses it to construct a Social Development Index (SDI) for the nineteenth century. The variables are extracted from their 1988 book. They are re-codified for convenience but the integrity of the database is preserved. Some variables available from Adelman and Morris (1988), their previous publications, and other posterior sources have been omitted because the alternative variables covering the same concept are preferable in terms of country classification and overall consistency of the database. Re-codification consisted of transferring letter codification (alphabetic order of categories) into numeric codification (categories sorted by ordinal numbers). This turned alphabetically coded variables into numerically coded variables, suitable for the intended statistical analysis. In addition, almost half of the variables have been renovated or updated.[6] Then we performed a principal components analysis.

3. III. 2. QUANTITATIVE METHODOLOGY

Why do I think that the principal components must convey information about the quality of the society? One could have gathered a bunch of variables which I think describe the quality of a society and calculate a simple average. However, not all variables contribute in the same measure to explain the dispersion in the data; some do capture more variance than others. The principal components analysis (PCA) is a sophisticated weighted average that gives more importance to those variables that deserve it in terms of variance explained. If there is something that makes the countries genuinely different, the principal components analysis will capture it.

[6] Please, refer to Chapter 2 for a detailed discussion of the Adelman and Morris (1988) database.

In short, the principal components analysis is a data-reduction technique. It aims at giving a description of the relationships between a set of variables in terms of a smaller set of *linear combinations* of these variables. These linear combinations are called Principal Components. The extent to which the relationships between the variables can be adequately described by a *small* set of new variables called Principal Components depends on the correlations between the original variables. The higher the correlation between the original variables, the smaller the number of Principal Components and, thus, the most effective the data reduction is.[7]

The latter characteristic of the Principal Component Analysis will turn out to be very helpful to us. It is well-known that in the social sciences – and especially in economics - many variables are highly correlated. Economic time-series tend to move together, and because of this it is often difficult to separate the effect that one variable has on another. However, here we are concerned with the parsimonious description of a high-dimensional object (many variables) into a small-dimensional one (one index), and as a result the Principal Components Analysis actually *takes advantage* of these high correlations. It is for this reason that we consider this technique particularly adequate for our purposes.

But the principal components analysis is not merely a data reduction exercise. The principal components are the underlying factors behind the variables, those factors that make them move together (covariate), and cannot be captured in any other way than in the abstract. The variables are just the reflection or result of those underlying factors that make them move together. And social capital is precisely this glue. This is why I think that a set of variables describing a society can help us capture the social capital behind.

3. III. 3. PRE-PCA TESTS: OPTIMISING THE VARIABLES

Prior to the principal components analysis (PCA) of the data, there are some statistical preliminary tests that are convenient to run. The first one of these preliminary tests is the Bartlett's test of sphericity. This test checks the validity of the whole exercise. Secondly, one might want to check for the validity of the inclusion of individual

[7] An intuitive explanation of the principal components technique can be found in appendix D.

variables. There are several ways to test the adequacy variable-by-variable. These are the square multiple correlation coefficient or R-squared of individual variables, a test of simple partial correlations, and, finally, Kaiser's measure of sampling adequacy[8]. This section goes through all these tests, prior to the principal components analysis.

The first preliminary test is the Bartlett's test of sphericity (Bartlett, 1950). This is a basic test of the significance of the principal components analysis[9]. It checks that the correlations matrix of all variables to be included in the analysis is not an identity matrix. Should that be the case, it would make no sense to run a principal components analysis because this type of analysis is only adequate for variables that are interrelated. The Bartlett's test of sphericity tests the null hypothesis that the correlation matrix is an identity matrix (all zeros except ones in the main diagonal) against the alternative hypothesis that the variables show statistically significant correlations amongst each other. If the null hypothesis is rejected, then the database has passed the first hurdle.

Given the socio-economic character of the variables in the database, one would be very surprised if the database would not pass this test. The principal components analysis has been chosen precisely because of the adequacy of this type of analysis to strongly interrelated variables.

The variables referring to the rate of adoption of new technologies in industry and agriculture respectively are perfectly correlated. The latter has been removed from the analysis in order to avoid multicolinearity problems. Therefore, there is one unique variable referring to the rate of adoption of new technologies, and this corresponds to both industry and agriculture.

As far as the adequacy of variables is concerned, all variables are in the database because there is a theoretical justification to include them. They are all related to or helping to reveal the quality of society in one way or another. However, statistical methods offer the possibility to put all variables through tougher tests of adequacy. First, the partial correlation between two variables is the correlation between the deviations from the mean. The simple partial correlations test consists of calculating partial correlations between all pairs of variables and checking whether these are strong

[8] These series of preliminary tests have been suggested by Wuensch (2005).
[9] Actually, the Bartlett's test of sphericity is designed to test the significance of any factor analysis, regardless of the extraction method.

enough for us to think they are contributing to the same underlying factors or components. Second, the squared multiple correlations (SMC) test calculates the R-squared of every variable with all the rest, and, therefore, is a more elaborated measure of adequacy variable-by-variable. For this reason, the first test of simple partial correlations is omitted. Still more sophisticated is the Kaiser-Meyer-Olkin measure of sampling adequacy (KMO). This measure is calculated as 1 minus the ratio between the sum of squares of all pair wise correlations (pair wise variances) and the sum of squares of all partial correlations of that variable with the rest[10]. This ratio is close to 1 when the pair wise correlations with all variables are high and the partial correlations are small, and close to 0 if the pair wise correlations with all other variables are small and the partial correlations high. It can be calculated both for individual variables and overall. For maximum representativeness in a principal components analysis one would want the KMO being close to 1, meaning that direct correlations with other variables are much more important than correlations of deviations from the mean. This is a way to measure the proportion of total variance explained by a particular variable with respect to the average. The Kaiser-Meyer-Olkin measure of sampling adequacy (KMO) is used as the main criterion of inclusion/exclusion of variables, although the SMC or R-squared is also calculated for reassurance/illustration purposes.

The Kaiser-Meyer-Olkin (KMO) measure of sampling adequacy can theoretically take values from minus infinity to 1. In practice, very rarely takes negative values, so it is regarded as running from 0 to 1. Small values mean that overall the variables have too little in common to warrant a good factor analysis. Heuristically, the following labels are given to values of KMO:

Less than 0.49 unacceptable
0.50 to 0.59 miserable
0.60 to 0.69 mediocre
0.70 to 0.79 middling
0.80 to 0.89 meritorious
0.90 to 1.00 marvelous,

as classified in the STATA User's Manual, based on Kaiser. This measure of sampling adequacy improves as the number of variables increases, the number of effective

[10] $KMO = 1 - \frac{\sum_{k \neq i} r_{ik}^2}{\sum_{k \neq i} pr_{ik}^2}$, where r is the correlation coefficient and pr is the partial correlation coefficient.

components decreases, the number of individuals increases, and the general level of correlation increases (Kaiser, 1970:405).

The KMO test is positive overall. The global rating for the adequacy of the sample of variables is 0.72 in a scale of 0 to 1. The KMO test also gives a mark for each one of the variables. A mark lower than 0.5 is suggesting the corresponding variable to be removed from the analysis. From the 32 variables in the first stage principal components analysis 4 are returning a mark lower than 0.5: Export growth group, population growth group, immigration group, and land concentration. Then, a second principal components analysis is performed.

In the second round (removing the 4 irrelevant variables) the overall KMO rate has improved from 0.72 to 0.77, and many more variables seem to fit better in this dimension. Also, 2 of them have become irrelevant: population per farm and population group. This is an invitation to perform a third principal components analysis round removing those two additional variables.

The third round with only 26 of the initial 32 variables is giving acceptable KMO marks for all variables. Therefore, it is the moment to stop. In addition, the overall KMO mark remains at 0.77 (same as last round). In total, we have removed 6 variables from the initial set plus 1 for being redundant. All the rest have been proven to have an empirical on top of theoretical justification to be there. It is interesting to note that the first principal component, characterised by the highest ranked variables, is becoming increasingly more important in terms of total variance captured as successive corrected rounds are performed. In the first round, the first principal component was capturing almost 45 percent of the variation in the data, in the second round it was capturing more than 50 percent, and in the third round it captures 54 percent.[11] The relative importance of the first principal component with respect to the rest is explored in more detail in the next section of this chapter.

As a robustness test for the list of variables eventually included in the PCA let us alter the KMO selection criteria. Before, it has been set at a minimum of 0.5 variable-by-variable, following the classification above. Let us move this barrier up to the next level, 0.6 mark or above variable-by-variable. This is a more strict selection criterion,

[11] Exact values are shown in the tables section at the end of the chapter.

and it is a standard one. Following this alternative criterion suggests dispensing with some extra variables after the first round: income growth, population per farmland, industrial wage change, agricultural wage change, population group, and form of land tenure, all of these scoring between 0.5 and 0.6, the latter not inclusive, [0.5, 0.6). Following this stricter criterion leaves a satisfying list of variables already in the second round. This has raised the KMO overall mark from 0.77 to 0.87, while leaving the significance level of the Barlett's sphericity test practically untouched. Thus, we shall stick to this criterion; all this keeping in mind that the overall KMO test is returning a global measure above 0.7 in all instances for the whole group of variables considered together. So, the successive rounds of variable-by-variable KMO test should be taken as a refinement of the selection exercise. As a result of the improved suitability, 22 variables are left from the original 32. The total variance explained by the first principal component alone rises from the initial 45 to almost 60 percent thanks to the optimisation of the choice of variables [12].

Now it is turn to turn into the new database, described in detail in chapter 2. The new database preserves the spirit of the original Adelman and Morris database but has been renovated and updated with the information available up to date. Therefore, it should be the preferred choice. The initial count of variables available for inclusion in the PCA goes up to 35, thanks to the updating work. A total of 14 variables are new, most of them substituting old ones. The same preliminary tests have now been run for the new database.

As before, the rate of adoption of new technologies in agriculture needs to be removed because of perfect correlation with the rate of adoption of new technologies in industry. Leaving just one of these two, no matter which one, will represent the rate of adoption of new technologies in general. This leaves the initial analysis of the new data with 34 variables to start with.

The old variables for economic role of the government (govt) and representativeness of political institutions (represent) have a clearly superior performance than the new ones (govrel and polity2) in being aligned with the components. This fact is an invitation for the reincorporation of previous alternatives. Let us first do the second round with the suppression of these two as the KMO test suggests. Then, we will move to

[12] See tables at the end of the chapter.

reincorporation of old adequate variables. In both cases, for the inclusion in the second round, the same restrictive criterion as above has been followed: All variables with a KMO mark below 0.6 have been suppressed.

The final result is the second round with the new database, having replaced govrel and polity2 by their old corresponding variables, govt and represent, due to the fact that they seem to perform better than the new ones. An initial survey has been made with the 34 initial variables, just replacing back these two. Then, all those scoring below 0.6 in the KMO test have been removed. This gives rise to the final selection of variables.

Results are indeed the best in terms of adequacy of the variables that contribute to explain the principal components. This is reflected in the fact that the overall KMO is above 0.9, which corresponds to the highest target for the KMO test, and described as "marvellous". Actually, once the less relevant variables have been suppressed, all 18 eventually included variables have an individual KMO mark around 0.9, as seen in the tables shown at the end of the chapter. This reflects the adequacy of the variables in capturing the underlying principal components. Actually, this selection of variables happens to return also the highest percentage of variance explained by the first principal component alone (above 65 percent), which reinforces the intuitive idea that a summary variable unlocked by the principal components is more than feasible. The principal components selection is explored in more detail in the next section.

Both datasets, an adaptation of the original by Morris and Adelman (1988) and the renewed one presented in this thesis, have been put to test for best PCA performance potential. Preliminary tests for optimal selection of variables reveal that the new dataset has a higher PCA potential, once cleaned up. However, some of the old variables have been recovered in the final selection, because they seem to capture elements not present in the new data. In all cases there are no problems of insufficient correlation amongst the variables, as shown by the Barlett's sphericity test.

3. III. 4. RESULTS: PRINCIPAL COMPONENTS ANALYSIS

The principal components are newly generated variables, obtained from linear combinations of the original. The first component extracts the linear combination of the variables that contains most of the variance. Successive orthogonal components explain

the variance that is left. Because every variable entering the analysis is startardised to have variance equal to 1, one should be interested in those outcoming principal components that have an eigenvalue greater than 1. This means they can explain more variance than any of the single variables can; the greater the eigenvalue, the larger the amount of variance explained. The sum of all eigenvalues has to be equal to the number of variables. This is why we are only potentially interested in those principal components with an eigenvalue greater than 1, although there are as many principal components (or dimensions) as variables.

The principal components are sorted by the associated eigenvalue: The first principal component is the one that has the highest eigenvalue and thus can explain the greatest amount of variance in the data, the second one is the one that has the second highest eigenvalue, and so on. Out of the 18 principal components extracted, there are only 3 with an eigenvalue greater than one. These three alone capture 80 percent of the total variation in the data. Actually, only one of the has a 2-digit percentage of explained variance. Not only this but, in this case, the first principal component alone explains more than 65 percent of the total variation in the data, making data reduction very effective.

Tables 3.10 and 3.11 show the total variance explained by the principal components in 1870 and 1890 respectively, sorted by the amount of variance in the data they are able to capture. Remarkably, the analysis reveals that the first principal component alone explains between 63 and 65 percent of the variation in the data for both years 1870 and 1890. The significance of this first principal component is also very clearly reflected in the scree plots (graphs 3.2 and 3.3 corresponding to 1870 and 1980 respectively). The figures for the eigenvalues abruptly decline after the first principal component. So we can take without fear the first principal component as the main underlying unobserved explanatory factor in the data.

How many components to retain? It can be observed both in the explained variance table and the scree plot that the first principal component is overwhelmingly the most representative one and effective in terms of data reduction. This is a quite clear case; nevertheless, it is always beneficial to double check with the theory. In a relatively recent methodological paper, Diana and Tommasi (2002) suggest to stop when the

associated eigenvalue is more than one and a half times the following one[13]. In this case, the ratio of the first to the second eigenvalue is in the order of 7, so we are far on the safe side.

The dominant principal component of the analysis of the renovated Adelman and Morris database can be interpreted as the level of socio-economic development. I have extracted a score for each and year I have data for. I named this variable SDI [YEAR], standing for social development index in a given year. Series for years 1870 and 1890 are presented here; though there is potential to expand the list of years to 1850 and 1910, with the necessary amount of time and resources. Most of the necessary data for the latter are already available, but the way lagged variables operated in the computations made the final index non-obtainable for the first and last periods, 1850 and 1914.

The first principal component score coefficients and scores for both years 1870 and 1890 are shown in the tables 3.12 and 3.13. The component scores coefficients shown in table 3.12 are the weights by which variables are multiplied to obtain the country scores. Full name and description of variables can be found in the appendix. Table 3.13 shows the first principal component scores obtained for every country and year in the sample. These are available for 23 countries (remarkably including China, country for which no data at this respect has ever been produced before, to the extent of my knowledge). In this way, we have a score for Argentina 1870, another for Argentina 1890, and so on, completing the list of 23 countries.[14] Two additional columns have been added to table 3.13 in order to monitor the evolution of the social development index over time. The third numerical column has been obtained by subtracting SDI 1870 from SDI 1890. The result is the change of the index in these two decades. The last column in table 3.13 indicates the sign of the change, either positive (increase) or negative (decrease).

Interestingly enough, practically all countries in the sample show an increase in social development for the period under study, 1870 to 1890. New Zealand and Japan are the countries that improved the most in the SDI (more than 1 standard deviation in 20 years), being the average improvement around half a standard deviation. The United

[13] In other words, max{j=1, …, k} Aj/Aj+i > 1.5 (Diana and Tommasi, 2002:80).
[14] List of countries in Appendix A.

Kingdom is the only country that appears to have lost some of its stock of social capital with respect to other countries. After the United Kingdom, but still exhibiting a positive sign of change, are China and France, in this order. There seems to the 2 groups of countries that have improved less than the rest of the sample, these being either Western European countries which departed from a privileged position and therefore do not have so much catching-up potential (UK, France, Belgium), or countries in a very poor stage of development which are not making much progress at that time yet (China, Egypt, India). These 2 groups are the countries that have lost or won less in terms of relative position in the ranking.

The two completely new series are depicted in figure 3.4. The Social Development Index for 1870 is positioned in the horizontal axis, while the 1890 counterpart lays on the vertical axis. In this way, we can see the change in the positioning of countries during the 20-year period in between. A diagonal 45-degree line has been drawn for ease of interpretation. All countries above the line improved their score in 1890 with respect to 1870. Countries below the line scored lower in 1890 than in 1870. Almost all countries managed to improve their score, as confirmed in table 3.13.

3. IV. LONG RUN INTER-TEMPORAL COMPARISONS: NINETEENTH AND TWENTIETH CENTURIES

Tabellini (2007) provides a framework for long run inter-temporal comparison. In his investigation for the European regions, he concludes that economic outcomes have their root in historically fundamented attitudes towards trust, respect for others and self-determination and, therefore, gives a quite deterministic view to economic development, in the direction of what Douglass North pointed at regarding historical determinism (North, 1981).

The newly developed Social Development Index for the late nineteenth century creates some room for a long run inter-temporal comparison of the quality of the society for countries all over the world, including both colonised and colonisers. The new SDI series for 1870 and 1890 can be contrasted to the contemporary measures of social capital. In particular SOCDEV for the early 1960's was constructed with a similar technique. Unfortunately, samples of countries for the nineteenth and the twentieth

century overlap thinly. This results in a small number of countries being in the two samples for this specific index.

Figures 3.5 and 3.6 depict the historical evolution of social developments over long periods of time. Figure 3.5 represents the change in scores over almost a century, from 1870 to 1960. Figure 2.6 depicts the change over a 70 year period, from 1890 to 1960. All countries in the sample have improved notably over these long periods of time.

We have just seen that tracing the change of the Social Development Index over the twentieth century is currently feasible for a small sample of countries. Now, what can we learn from the relationship of the nineteenth century SDI with other twentieth century indicators of social capital? At this point, it turns useful to bring into the analysis the two most popular contemporary alternatives, namely trust and civic engagement. In particular, are there any patterns in which these two contemporary variables proceeding from surveys relate to the nineteenth century newly constructed estimates?

Starting with the most recent first, the relationship between TRUST and CIVIC is illustrated by means of the overlay scatter plot (i. e. overlapping to scatter plots). The scatter plots should be read in the same way described for the twentieth century analysis in section 3.II. In figures 3.7 to 3.12 the light coloured dots and lines represent the relation between SDI and TRUST, while the dark dots and lines represent the relation between SDI and CIVIC.

Figure 3.7 shows quite different from its twentieth century counterpart studied in section 3.II (figure 3.2). In the first place, countries are more widely spread over the social development index range and less over the vertical axis. This indicates convergence from a wide range of social development positions in the nineteenth century to a more equalised level at the end of the twentieth century. Late nineteenth century results are similar (see figure 3.8 for 1890). Secondly, means for contemporary variables are reversed. Now the mean of TRUST is higher than the mean of CIVIC, both when contrasted to past SDI. This is true both for 1870 and 1890. Also, civic engagement appears to be more stable or equalised across countries than trust. This means that all countries in the display a relatively similar level of contemporary civic engagement, no matter where they were standing in terms of social development in the

past. However, this is not true for trust. There seems to be a pattern in the distribution of trust across countries, depending what was their departure point in terms of social development in the past. Let us have a closer look at this phenomenon.

Figures 3.9 and 3.10 show the same set of data with fitted lines for 1870-nowadays and 1890 nowadays historical evolutions. In both cases the fitting method was the lowess method, with fifty percent of the points fitted in three iterations (same as in section 3.II graphs). Here we can observe that trust is more volatile across countries than civic engagement is, and has a tendency to be less equalised. This fact stands clear from both 1870-nowadays and 1890-nowadays fit lines. Again, India stands as the most paradigmatic outlier in the sample, showing a spectacular social evolution in the course of the twentieth century. New outliers revealed by the nineteenth century analysis are Norway in the good side, and Brazil, Australia, and France in the down side. We would not have expected this deceiving result from France or Australia, even with the more than one century's perspective. But be aware that we have only very recently realised that India had a big potential for economic growth, which has recently being spectacularly coming out. This was not obvious just twenty years ago. So, we are afraid one could be a catastrophist when auguring growth prospects for France and Australia if one is to judge by the social evolution indications[15].

The tendency of civic engagement levels to be equalised across countries regardless of their past is confirmed in figures 3.11 and 3.12. These present fit lines with ninety-five percent confidence intervals. Quadratic and cubic regression prediction lines were used respectively, according to which method fitted the data best. Again 1870-nowadays and 1890-nowadays analyses show similar results, the main ones being: 1) civic engagement tends to be similar for all countries in the sample, regardless of where they were standing in terms of social development in the past and, 2) parabolic layout of trust observations.

The contemporary civic engagement levels appear to be very similar for almost every country in the sample, regardless of what was the level of social development in the late nineteenth century. Second, open ends mean that social extremes seem to be more unpredictable. This happens by construction of the confidence intervals. Extremes tend to be more unpredictable, since we only have data either from the right or from the left,

[15] Implications for per capita income are tested in the next chapter (Chapter 4).

but not from both sides. Still, this phenomenon is especially acute for the bottom tale of the sample. Countries with very low levels of social development in the late nineteenth century have proved to unfold in all directions: They might evolve into a miracle (India) or turn down to a catastrophe (Brazil).

The 'parabolic layout of trust points' means that the cross-country study of the evolution of trust contrasted with social development in the nineteenth century reveals a parabolic layout. Even allowing for a higher level polynomial would the prediction line turn out to be quasi-parabolic (see figure 3.12). The relevance of the quadratic term can be tested econometrically. It can be shown that the square of social development in a trust regression is significant at the standard 5 percent level. In other words, countries in the middle of the spectrum have improved the most with respect to their nineteenth century position in the SDI ranking. I interpret this parabola as the combination of two phenomena: the unpredictable direction that the worst scored countries will follow (see previous paragraph), combined with the Abramovitz hypothesis of 'falling behind' for the best historically positioned countries (Abramovitz, 1986). In this way, countries in the middle of the spectrum have the highest predictable prospects for catching up.

A considerable historical perspective is added to the analysis. We can observe a tendency to persistence of the social indicators. So, there is an element of North's hypothesis on path dependency, as suggested by Tabellini (2007). However, outliers depart from the trend, doing nothing but confirm that the results are historically consistent with future economic growth trajectories. This is the case of India, which shows exceptionally high values in the social development index or Brazil, whose scores are deceivingly poor. A striking characteristic is the finding that some socially well located countries in the nineteenth century show to be loosing their relative position at the end of the twentieth century. This preoccupating phenomenon, which surprises as counterintuitive, needs a more detailed consideration. One needs to keep in mind that the comparison is done with different scales. But, as a first approximation, this dramatic finding is nothing but the proof of what Putnam was pointing at in his 2000 book *Bowling Alone*, detailing the weakening of social values in the North American society. There seems to be a tendency for Western European countries to fall into this group.

3. V. CONCLUSION

We presented the first international historical estimates for social capital. Two new series of a Social Development Index (SDI) become available: one for 1870 and one for 1890. We showed a new way of looking at social evolution. Together with some other contemporary measurement attempts, the new series allow monitoring the evolution of a social development index over time.

As a conclusion, one can say that it is true that social attitudes are correlated to economic performance (Tabellini, 2007), but it is not true that the former do not change. The inter-temporal comparison of the various social capital proxies shows that cultural revolutions are possible and are indeed observed in the data if one allows for a sufficient time span. The present study looks at the change in the world country ranking of social capital proxies for a period of 70 to 90 years, and the result is that cultural miracles (as growth miracles) exist. When all regions in the world are included instead of only Europe, differentiated patterns of social development trajectories can be pinned down.

Thus, North's hypothesis of path dependency is modified. Practically all countries in the sample show an increase in social development during the intermediate period (1870 to 1890), and all of them reveal a very significant improvement over the twentieth century. We find some outstanding performers in social improvement, defeating path dependence, and also detect some failure stories. In both cases, the social development trajectories seem to be historically consistent with their subsequent economic growth experiences.

Europe's relative position with respect to the rest of the World varies. Scandinavian countries are absolute leaders on trust, while they were in the centre of the social development spectrum more that 100 years ago. Meanwhile, some core Western European countries like France or the United Kingdom, who were World leaders once, seem to have lost their privileged positions during the course of the twentieth century.

Finally, different social capital measurement alternatives exhibit different patterns, suggesting that they are simply capturing different aspects. We find statistically significant non-linear relationships between them. In particular, trust describes a

parabolic layout with respect to our Social Development Index, and civic engagement stands as surprisingly even across countries.

TABLES AND FIGURES FOR SECTION 3.II - CONTEMPORARY COMPARISON OF ALTERNATIVES

Table 3.1 – Three alternative measures to monitor social capital

COUNTRY	SOC DEV	TRUST AM	CIVIC KK	COUNTRY	SOC DEV	TRUST AM	CIVIC KK
Afganistan	-1,02	.	.	Lesotho	.	.	.
Algeria	0,18	.	.	Liberia	-1,01	.	.
Angola	.	.	.	Libya	-0,68	.	.
Argentina	1,91	27.0	39.50	Lithuania	.	22	.
Armenia	.	25	.	Luxembourg	.	.	.
Australia	.	47.8	38.27	Madagascar	-1,31	.	.
Austria	.	31.8	41.45	Malawi	-1,57	.	.
Azerbaijan	.	21	.	Malaysia	.	.	.
Bahamas, The	.	.	.	Mali	.	.	.
Bahrain	.	.	.	Malta	.	.	.
Bangladesh	.	21	.	Mauritania	.	.	.
Barbados	.	.	.	Mauritius	.	.	.
Belarus	.	24	.	Mexico	0,75	17.7	34.55
Belgium	.	30.2	38.08	Moldova	.	22	.
Benin	-1,54	.	.	Morocco	-0,57	.	.
Bolivia	-0,35	.	.	Mozambique	.	.	.
Botswana	.	.	.	Myanmar (Burma)	-0,41	.	.
Brazil	0,79	6.7	37.58	Nepal	-1,36	.	.
Bulgaria	.	30.4	.	Netherlands	.	46.2	38.36
Burkina Faso	.	.	.	New Zealand	.	.	.
Burundi	.	.	.	Nicaragua	0,88	.	.
Cambodia	-0,55	.	.	Níger	-1,86	.	.
Cameroon	-1,34	.	.	Nigeria	-0,91	22.9	39.19
Canada	.	49.6	39.74	Norway	.	61.2	40.75
Cape Verde	.	.	.	Oman	.	.	.
Central African Rep.	.	.	.	Pakistan	-0,08	.	.
Chad	-1,70	.	.	Panama	0,84	.	.
Chile	1,39	22.7	36.80	Papua New Guinea	.	.	.
China	.	.	.	Paraguay	0,97	.	.
Colombia	0,66	10	.	Peru	0,68	5	.
	SOC DEV	TRUST AM	CIVIC KK	COUNTRY	SOC DEV	TRUST AM	CIVIC KK

Comoros	.	.	.	Philippines	0,56	6	.
Congo	.	.	.	Poland	.	34.5	.
Costa Rica	0,78	.	.	Portugal	.	21.4	36.89
Cote d'Ivoire	-0,98	.	.	Romania	.	16.1	.
Croatia	.	25	.	Russia	.	24	.
Cyprus	1,08	.	.	Rwanda	.	.	.
Czech Republic	.	30	.	Saudi Arabia	.	.	.
Denmark	.	56.0	40.34	Senegal	-0,52	.	.
Dominica	.	.	.	Seychelles	.	.	.
Dominican Rep.	0,81	26	.	Sierra Leone	-1,39	.	.
Ecuador	0,54	.	.	Singapore	.	.	.
Egypt	0,73	.	.	Slovakia	.	23	.
El Salvador	0,71	.	.	Slovenia	.	16	.
Estonia	.	22	.	Solomon Islands	.	.	.
Ethiopia	-0,99	.	.	Somalia	-1,35	.	.
Fiji	.	.	.	South Africa	0,62	30.5	36.99
Finland	.	57.2	40.64	Spain	.	34.5	38.75
France	.	24.8	36.26	Sri Lanka	0,35	.	.
Gabon	-0,83	.	.	St.Lucia	.	.	.
Gambia	.	.	.	St.Vincent&Grens.	.	.	.
Georgia	.	23	.	Sudan	-0,64	.	.
Germany	.	29.8	39.83	Suriname	0,54	.	.
Ghana	-0,01	23	.	Swaziland	.	.	.
Greece	1,47	.	.	Sweden	.	57.1	41.57
Grenada	.	.	.	Switzerland	.	43.2	40.89
Guatemala	0,35	.	.	Syria	0,57	.	.
Guinea	-1,47	.	.	Taiwan	1,05	42	.
Guinea-Bissau	.	.	.	Tanzania	-1,22	.	.
Guyana	.	.	.	Thailand	0,50	.	.
Haiti	.	.	.	Togo	.	.	.
Honduras	0,26	.	.	Tonga	.	.	.
Hong Kong	.	.	.	Trinidad & Tobago	1,15	.	.
Hungary	.	24.6	.	Tunisia	-0,18	.	.
Iceland	.	41.6	41.07	Turkey	0,88	10.0	42.43
	SOC DEV	**TRUST AM**	**CIVIC KK**	**COUNTRY**	**SOC DEV**	**TRUST AM**	**CIVIC KK**
India	-0,28	34.3	42.65	Uganda	-1,22	.	.

Indonesia	-0,40	.	.	Ukraine	.	31	.
Iran, I.R. of	0,09	.	.	United Arab Emirates	.	.	.
Iraq	-0,03	.	.	United Kingdom	.	44.4	40.07
Ireland	.	40.2	37.51	United Status	.	45.4	40.55
Israel	1,77	.	.	Uruguay	1,59	22	.
Italy	.	26.3	41.23	Vanuatu	.	.	.
Jamaica	1,06	.	.	Venezuela	1,37	14	.
Japan	1,63	40.8	41.79	Vietnam, South	-0,49	.	.
Jordan	0,16	.	.	Western Samoa	.	.	.
Kenya	-0,53	.	.	Yemen, N.Arab	-1,35	.	.
Korea	0,85	38	.	Yugoslavia	.	31	.
Kuwait	.	38.0	39.64	Zaire	.	.	.
Laos	-1,06	.	.	Zambia	-0,89	.	.
Latvia	.	25	.	Zimbabwe	0,14	.	.
Lebanon	1,44	.	.				

Table 3.2 - Pearson Correlation Matrix of Social Capital Alternatives

		socdev 1960	trustam 1990	civickk 1990
socdev 1960	Pearson Correlation	1	,090	-,089
	Tail probability		,723	,820
	Number of cases	75	18	9
trustam 1990	Pearson Correlation	,090	1	,387(*)
	Tail probability	,723		,038
	Number of cases	18	58	29
civickk 1990	Pearson Correlation	-,089	,387(*)	1
	Tail probability	,820	,038	
	Number of cases	9	29	29

* Correlation is significant at the 0.05 level (2-tailed).

The correlation matrix shows that there is no strong linear relationship between these three variables.

Figure 3.1 - 3D Scatter Plots of Social Capital Alternatives. Spikes to the Floor and Centroid respectively

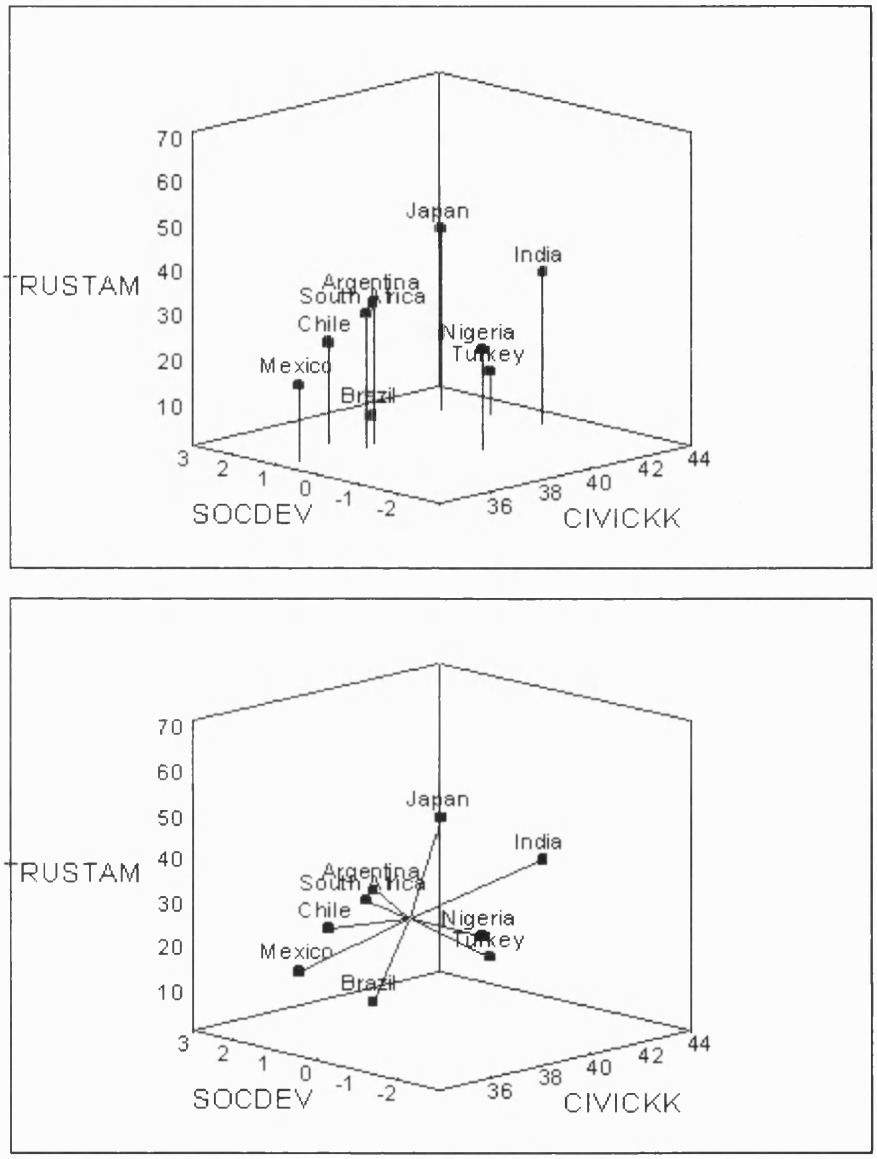
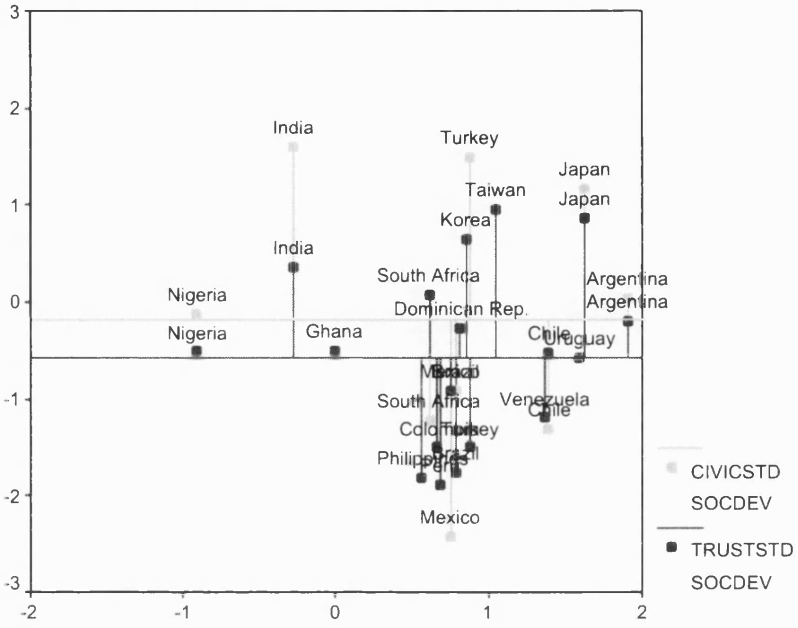
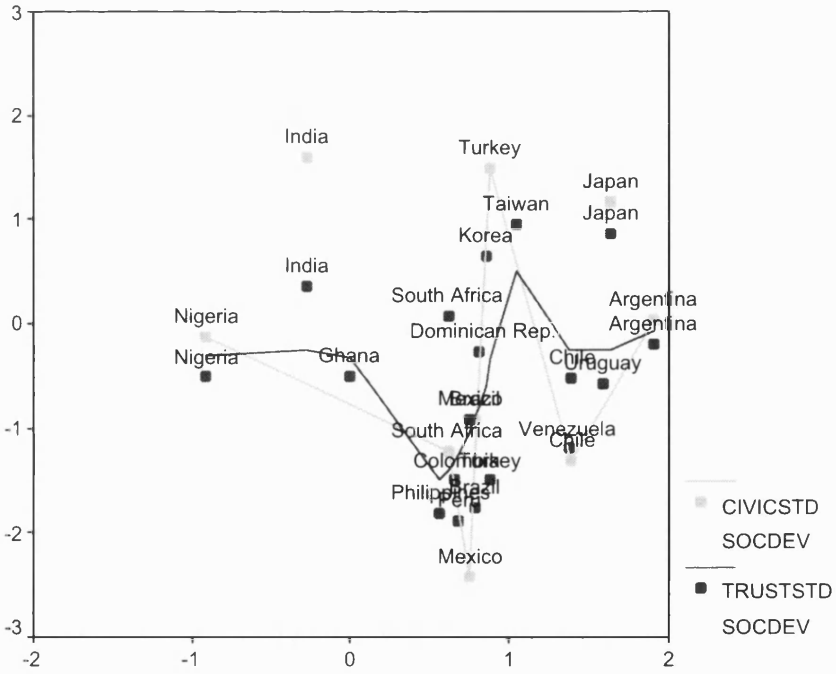


Figure 3.2 - Historical Evolution, from 1960 (Soc Dev) to Surveys: Overlay Scatter Plot



SOCDEV in the horizontal axis, CIVICSTD (light) and TRUSTSTD (dark) in the vertical axis. Spikes to reference line for each pair. Reference lines are mean of Y

Figure 3.3 - Overlay Scatter Plot with Fit Line



SOCDEV in the horizontal axis, CIVICSTD (light) and TRUSTSTD (dark) in the vertical axis. Fit Method: Lowess. 50% of points fitted with 3 iterations.

TABLES FOR 3.III.3 - PRE-PCA TESTS: OPTIMISING THE VARIABLES

KMO and Bartlett's Test

Kaiser-Meyer-Olkin Measure of Sampling Adequacy.		,716
Bartlett's Test of Sphericity	Approx. Chi-Square	1684,231
	df	496
	Sig.	,000

Table 3.3 – First round variables, after excluding rate of adoption of new technologies in agriculture

KMO and Bartlett's Test

Kaiser-Meyer-Olkin Measure of Sampling Adequacy.		,769
Bartlett's Test of Sphericity	Approx. Chi-Square	1419,284
	df	325
	Sig.	,000

Table 3.4 – Third (and last) round variables, after excluding rate of adoption of new technologies in agriculture, population per farmland, export growth group, the demographic variables (population group, population growth group and immigration group) and land concentration.

KMO and Bartlett's Test

Kaiser-Meyer-Olkin Measure of Sampling Adequacy.		,869
Bartlett's Test of Sphericity	Approx. Chi-Square	1258,497
	df	231
	Sig.	,000

Table 3.5 – Second (and last) round variables following a stricter criterion of variable-by-variable KMO selection (mark greater or equal than 0.6) for every variable. This has raised the overall KMO measure of sampling adequacy from 0.77 to 0.87, while keeping the significance of the sphericity test practically untouched.

KMO and Bartlett's Test

Kaiser-Meyer-Olkin Measure of Sampling Adequacy.		,909
Bartlett's Test of Sphericity	Approx. Chi-Square	1093,367
	df	153
	Sig.	,000

Table 3.6 – Final selection of variables

Table 3.7 - Squared Multiple Correlations (SMC) of variables with all other variables

Variables	Original Data		New Data	
	SMC initial variables	SMC final variables, after KMO test	SMC initial variables	SMC final variables, after KMO test
income	.9841	.9399	.9790	.9270
income growth	.9330	-	.8930	-
technique in industry	.9702	.9290	.9634	.9258
adoption of new techniques	.9349	.8956	.9776	.8501
technique in agriculture	.9714	.9152	.9799	.9029
agricultural labour	.9700	.9043	.9820	.8839
population per farmland	.8578	-	.9160	-
inland transportation	.9581	.9046	.9346	.8739
transportation, growth	.8049	.6294	.9151	.5736
export growth group	.7242	-	-	-
trade	-	-	.9618	-
shift in export structure	.8594	.7928	.8747	.6935
industrial wage change	.8491	-	.8840	-
agricultural wage change	.7164	-	.8710	-
wages	-	-	.8971	-
population	.8951	-	.9654	-
population growth	.8988	-	.9506	-
immigration	.8980	-	.8947	-
migration to pop. growth	-	-	.9050	-
literacy	.9748	.9309	.9879	.8827
primary education growth	.8011	.7330	.9401	-
form of land tenure	.9287	-	.9348	-
land concentration	.7499	-	.8628	-
land adoption	.9302	.7819	.9329	.7689
urbanisation	.9321	.8354	.9361	.7707
entrepreneurship	.9584	.9188	.9841	.8941
role of government	.7112	.6152	.6427	-
socio-politics	.9145	.8743	.9741	.8555
representativeness	.9488	.9042	-	.8479

polity 2	-	-	.9659	-
political stability	.9595	.8665	.9579	-
foreign dependency	.9161	.7965	.9400	-
colonial status	.9414	.8742	.9639	.7587
market development	.9810	.9511	.9922	.9318
market development growth	.9989	.9975	.9988	.9972
mkt develpt growth, lagged	.9989	.9972	.9987	.9969
Number of variables	**32**	**22**	**34**	**18**
KMO sampling adequacy	**.7161**	**.8692**	**.6026**	**.9086**

The SMC shows that all variables are strongly correlated to one another. However the sampling adequacy test reveals that not all of them are so relevant for the components analysis.

Table 3.8 – Kaiser-Meyer-Olkin (KMO) sample adequacy scores variable-by-variable

Variables	Original Data		New Data	
	KMO initial variables	KMO final variables	KMO initial variables	KMO final variables
income	.6985	.8589	.8194	.8938
income growth	.5393	-	.5443	-
technique in industry	.7590	.8866	.8517	.8856
adoption of new techniques	.8853	.8781	.6678	.9126
technique in agriculture	.7950	.9330	.6745	.9389
agricultural labour	.6579	.8512	.6578	.8889
population per farmland	.5702	-	.3341	-
inland transportation	.8314	.8674	.8629	.9161
transportation, growth	.7421	.8454	.4820	.8683
export growth group	.4345	-	-	-
trade	-	-	.1977	-
shift in export structure	.7839	.8170	.6892	.9214
industrial wage change	.5994	-	.5535	-
agricultural wage change	.5819	-	.2956	-
wages	-	-	.2857	-
population	.5216	-	.4677	-
population growth	.4177	-	.4521	-
immigration	.4658	-	.2399	-
migration to pop. growth	-	-	.4671	-
literacy	.8090	.8931	.7048	.9332
primary education growth	.6685	.7167	.3733	-
form of land tenure	.5909	-	.5719	-
land concentration	.2078	-	.1454	-
land adoption	.7264	.9337	.6294	.9288
urbanisation	.6957	.8547	.6394	.9138
entrepreneurship	.9029	.9515	.6509	.9444
role of government	.6380	.6545	.3797	-
socio-politics	.8491	.9230	.6929	.9372
representativeness	.8283	.8854	-	.9287
polity 2	-	-	.4022	-

political stability	.6373	.8033	.6323	-
foreign dependency	.7443	.8602	.6904	-
colonial status	.6276	.7379	.5800	.8168
market development	.7861	.9014	.6380	.9322
market development growth	.7401	.8683	.7381	.8713
mkt develpt growth, lagged	.7096	.8651	.7573	.8675
Number of variables	**32**	**22**	**34**	**18**
Overall sampling adequacy	**.7161**	**.8692**	**.6026**	**.9086**

The score differences between the initial and the final selection of variables reflect the improvement in sampling adequacy, in a scale from 0 to 1.

TABLES AND FIGURES FOR 3.III.4 - RESULTS: PRINCIPAL COMPONENTS ANALYSIS

Table 3.9 - Total Variance Explained

Principal components/correlation Number of obs = 46
Number of comp. = 18
Trace = 18
Rotation: (unrotated = principal) Rho = 1.0000

	Initial Eigenvalues			Extraction Sums of Squared Loadings		
Component	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	11,723	65,126	65,126	11,723	65,126	65,126
2	1,648	9,158	74,284	1,648	9,158	74,284
3	1,068	5,932	80,216	1,068	5,932	80,216
4	,889	4,936	85,152			
5	,664	3,691	88,843			
6	,453	2,516	91,359			
7	,303	1,683	93,042			
8	,287	1,593	94,635			
9	,209	1,161	95,796			
10	,183	1,017	96,813			
11	,142	,788	97,601			
12	,104	,580	98,181			
13	,087	,482	98,663			
14	,084	,467	99,130			
15	,069	,382	99,512			
16	,051	,284	99,796			
17	,035	,196	99,992			
18	,001	,008	100,000			

Codes of the variables included: income indutech indtecgr agritech agrilgrp intransp transpgr shiftx lit landadop urbani entrep sociopol represen colstat mktdev mktdevgr mktdvgrl Extraction Method: Principal Component Analysis.

This is the second round of the new database, but having replaced govrel and polity2 with the old ones govt and represen to start with, given that they give better results than the new ones. As a result, govt should be removed in any case, but represen stays in the second round. Re-introducing represen makes polstabi and foreignd go; in addition, transpgr cannot be removed.

Graph 3.1 - Scree Plot

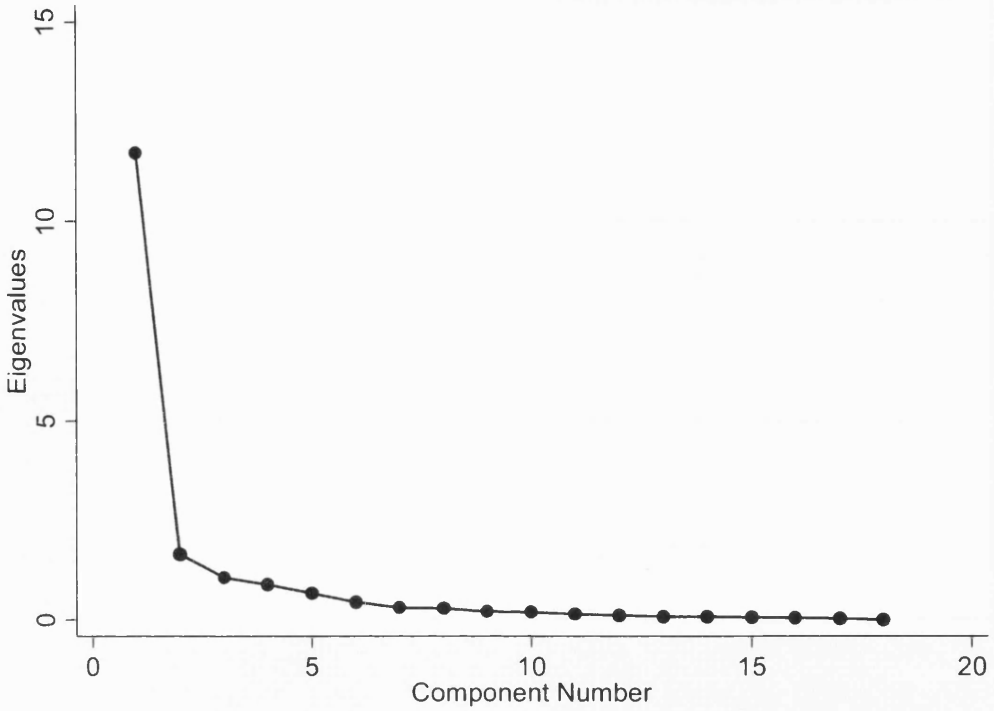


Table 3.10 - Total Variance Explained for 1870

Component	Initial Eigenvalues			Extraction Sums of Squared Loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	11,873	65,961	65,961	11,873	65,961	65,961
2	1,683	9,348	75,308	1,683	9,348	75,308
3	1,259	6,992	82,300	1,259	6,992	82,300
4	,703	3,904	86,204			
5	,693	3,851	90,055			
6	,549	3,053	93,108			
7	,306	1,698	94,806			
8	,207	1,150	95,956			
9	,192	1,064	97,021			
10	,136	,756	97,777			
11	,131	,725	98,502			
12	,084	,468	98,970			
13	,061	,340	99,310			
14	,053	,295	99,605			
15	,033	,185	99,790			
16	,024	,133	99,923			
17	,014	,076	100,000			
18	8,04E-005	,000	100,000			

Codes of the variables included: income indutech indtecgr agritech agrilgrp intransp transpgr shiftx lit landadop urbani entrep sociopol represen colstat mktdev mktdevgr mktdvgrl
Extraction Method: Principal Component Analysis.
Only cases for which year = 1870 are used in the analysis phase.

Graph 3.2 - Scree Plot for 1870

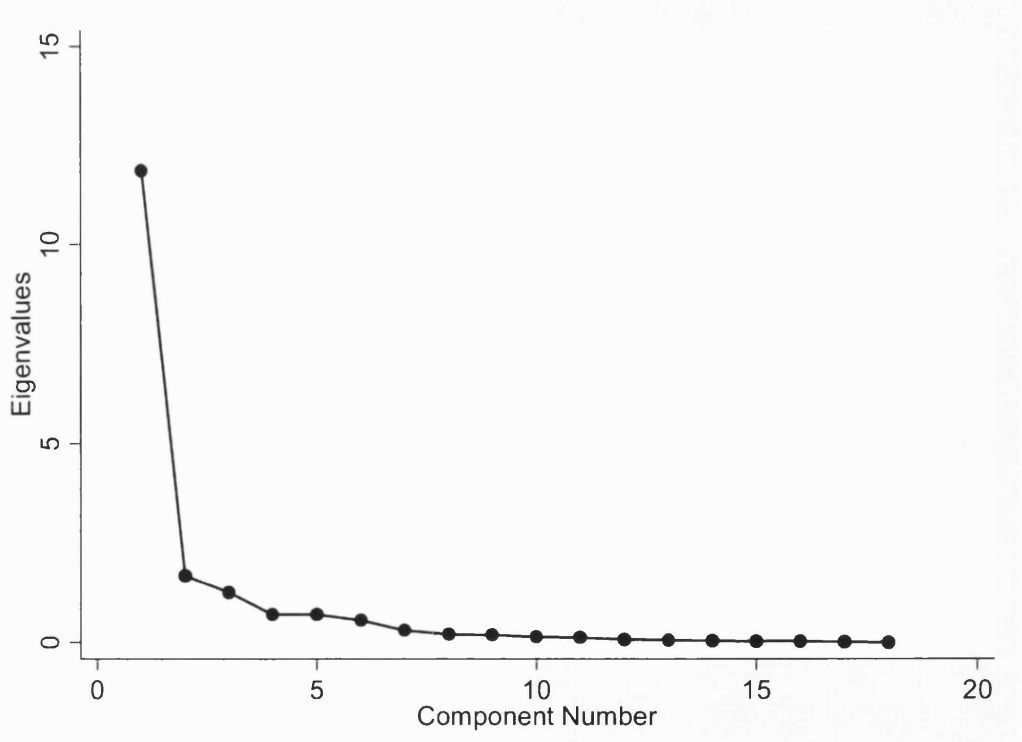


Table 3.11 - Total Variance Explained for 1890

Component	Initial Eigenvalues			Extraction Sums of Squared Loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	11,446	63,587	63,587	11,446	63,587	63,587
2	2,038	11,320	74,907	2,038	11,320	74,907
3	1,279	7,105	82,012	1,279	7,105	82,012
4	,915	5,086	87,098			
5	,587	3,258	90,356			
6	,450	2,500	92,856			
7	,326	1,812	94,668			
8	,246	1,366	96,034			
9	,187	1,041	97,075			
10	,138	,765	97,840			
11	,115	,639	98,478			
12	,082	,453	98,932			
13	,075	,417	99,348			
14	,057	,315	99,664			
15	,047	,259	99,923			
16	,011	,061	99,984			
17	,003	,015	99,999			
18	,000	,001	100,000			

Codes of the variables included: income indutech indtecgr agritech agrilgrp intransp transpgr shiftx lit landadop urbani entrep sociopol represen colstat mktdev mktdevgr mktdvgrl
Extraction Method: Principal Component Analysis.
Only cases for which year = 1890 are used in the analysis phase.

Graph 3.3 - Scree Plot for 1890

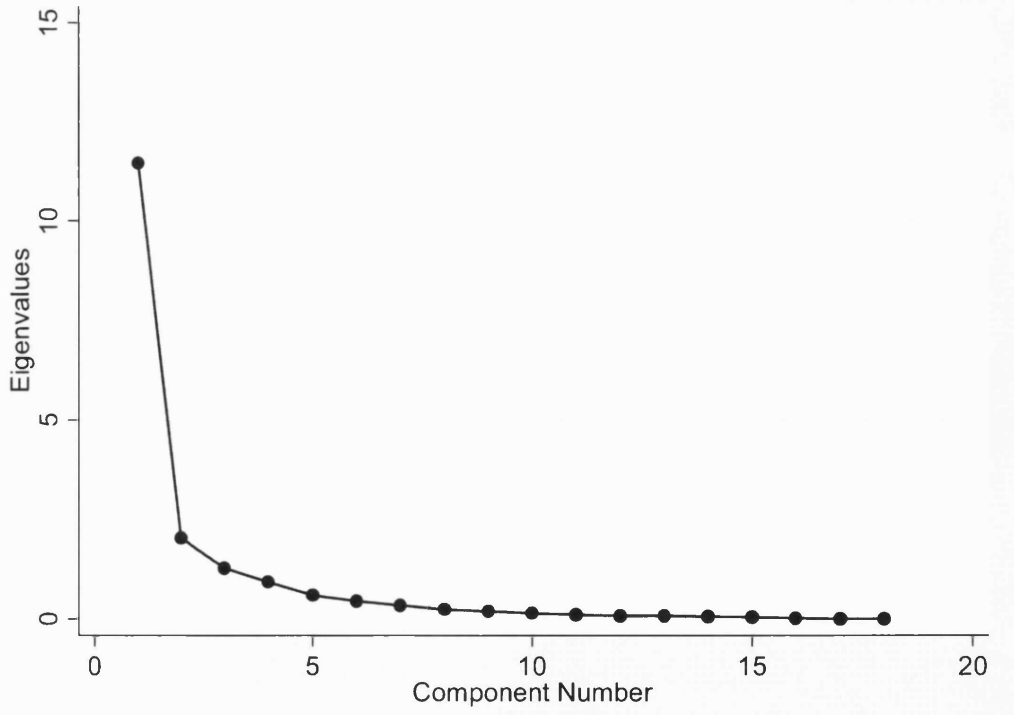


Table 3.12 - First Principal Component Score Coefficients (Loadings)

Variable	(1) Joint score	(2) Score 1870	(3) Score 1890
Income	.2463	.2363	.2539
Technique in industry	.2603	.2605	.2583
Adoption of new techniques	.2368	.2492	.2230
Technique in agriculture	.2650	.2630	.2666
Agricultural labour	-.2229	-.2181	-.2282
Inland transportation	.2445	.2319	.2577
Transportation, growth	.1744	.1833	.1496
Shift in export structure	.1897	.1907	.1864
Literacy	.2467	.2290	.2667
Land adoption	.2217	.2208	.2362
Urbanisation	.2036	.2041	.1953
Entrepreneurship	.2537	.2574	.2531
Socio-politics	.2509	.2500	.2515
Representativeness	.2529	.2468	.2565
Colonial status	.1731	.1818	.1770
Market development	.2750	.2721	.2787
Market development growth	.2486	.2601	.2338
Mkt develpt growth, lagged	.2435	.2578	.2247

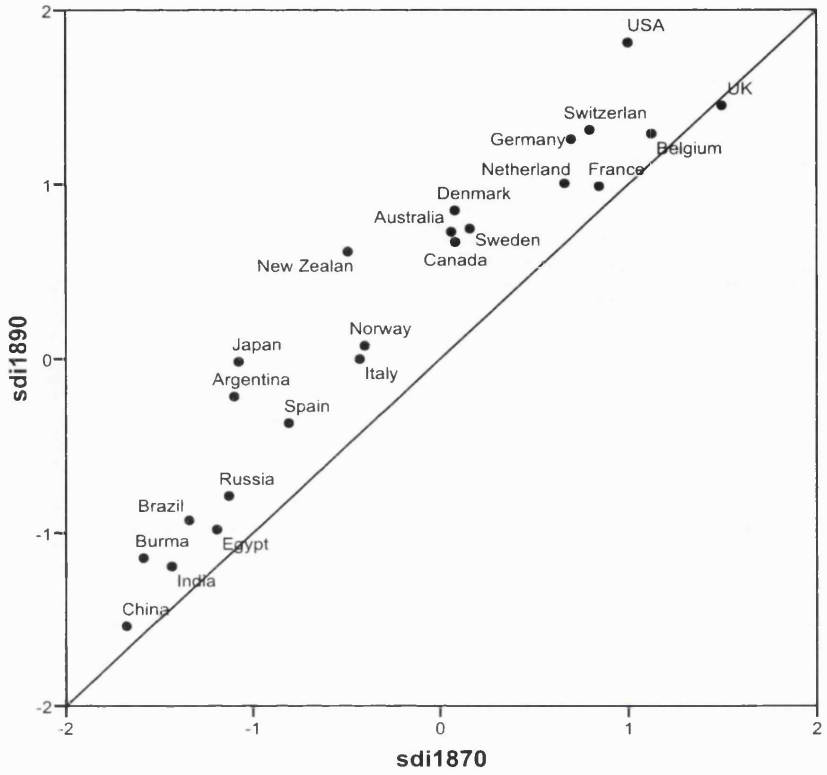
Notes: The weight given to each variable is determined by the eigenvectors of the correlation matrix of all variables. Columns in the table are the eigenvectors associated to the first principal component. (1) Joint score coefficients; common weights across periods. This is the preferred option. (2) Weights corresponding to 1870 only; (3) weights corresponding to 1890 only.
Normalisation: Sum of Squares column =1

Table 3.13 – First Principal Component Scores

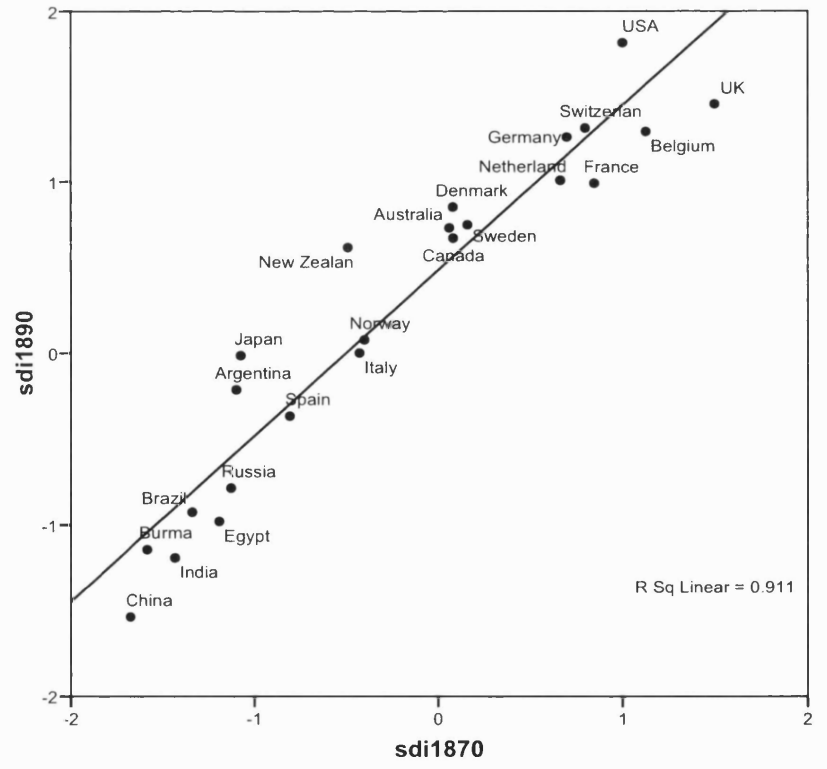
Country	SDI 1870	SDI 1890	Change SDI 1890 –SDI 1870	Sign of Change
Argentina	-1.099	-0.215	0.885	+
Australia	0.060	0.730	0.669	+
Belgium	1.125	1.292	0.168	+
Brazil	-1.339	-0.926	0.413	+
Burma	-1.581	-1.144	0.437	+
Canada	0.081	0.669	0.588	+
China	-1.673	-1.538	0.135	+
Denmark	0.079	0.851	0.772	+
Egypt	-1.192	-0.980	0.212	+
France	0.846	0.990	0.145	+
Germany	0.695	1.261	0.566	+
India	-1.433	-1.192	0.242	+
Italy	-0.429	-0.000	0.429	+
Japan	-1.075	-0.015	1.060	+
Netherlands	0.660	1.007	0.348	+
New Zealand	-0.492	0.615	1.107	+
Norway	-0.402	0.076	0.478	+
Russia	-1.128	-0.786	0.341	+
Spain	-0.809	-0.367	0.442	+
Sweden	0.158	0.747	0.589	+
Switzerland	0.796	1.314	0.518	+
United Kingdom	1.496	1.455	-0.042	-
United States	0.999	1.814	0.815	+

Note: The scores presented here have been standardised to have mean 0 and standard deviation 1. This results in a maximum observed score of 1.814 and a minimum of -1.673. To recover the unstandardised scores, multiply by 3.424. Italy score -0.000 is marked with the negative sign because for more than 4 significative ciphers, the number is negative.

Figure 3.4 - Scatter Plot for the New Social Development Index: 1870 against 1890

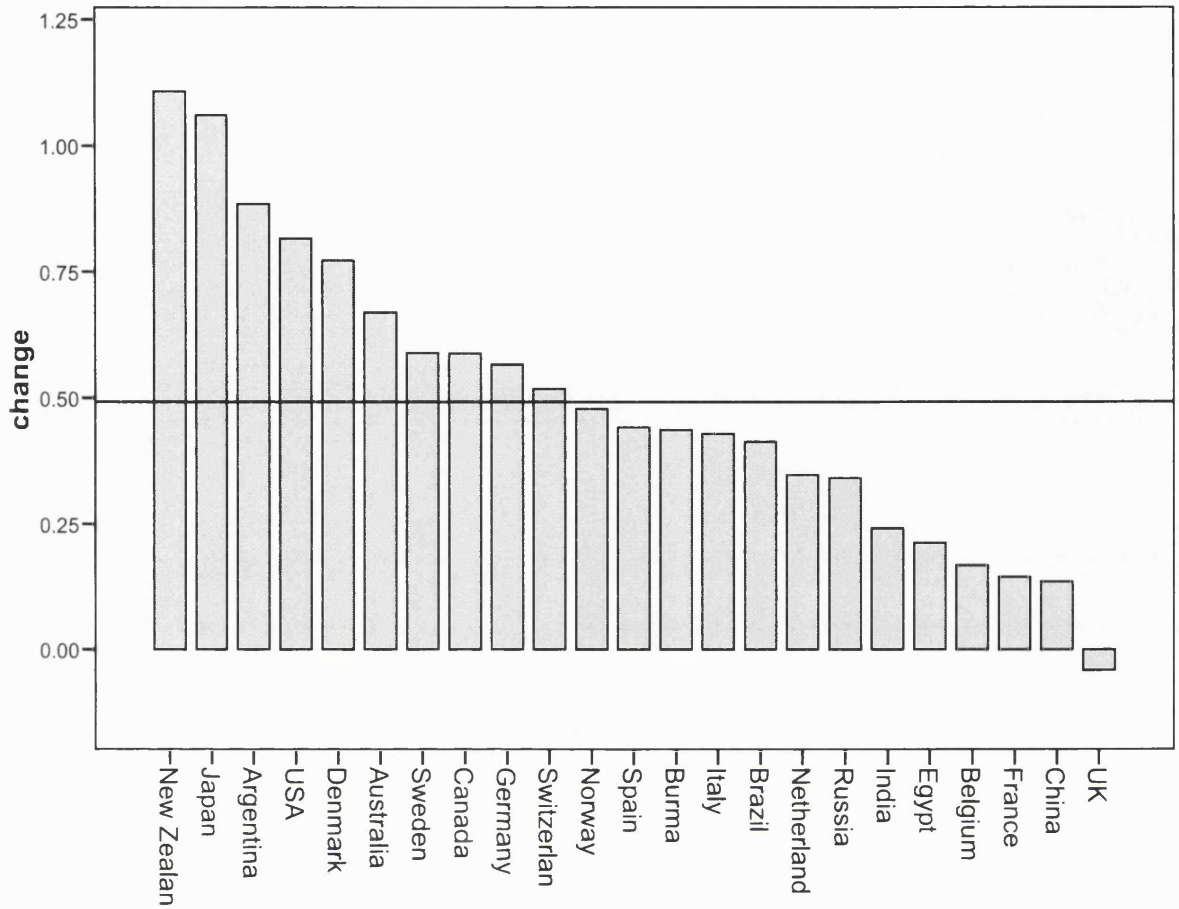


SDI 1870 in the horizontal axis. SDI 1890 in the vertical axis. Almost all countries lie above the 45-degree line. This means all countries except UK improved over this period.



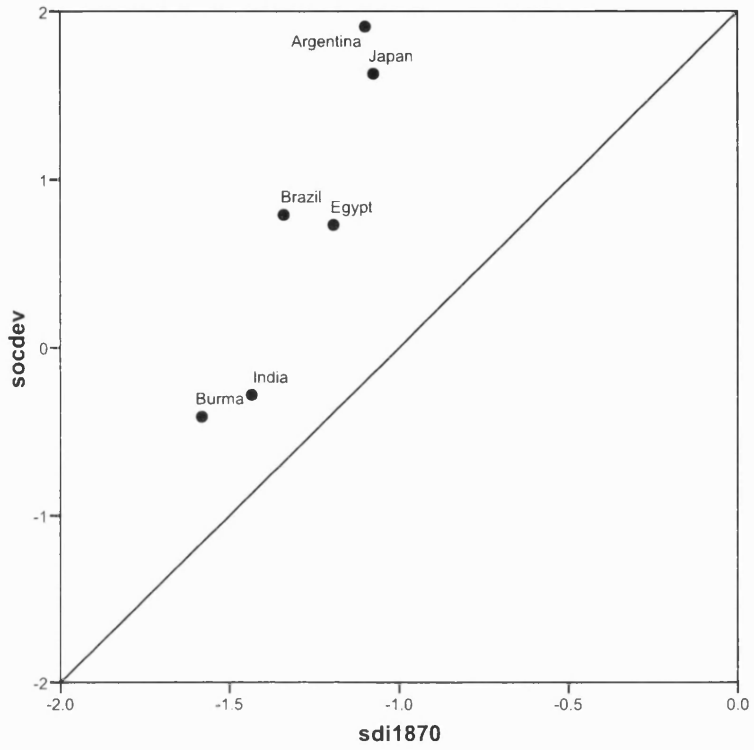
Scores with fit line at average standard deviation. Scores have improved an average of around half standard deviation in 2 decades.

Graph 3.4 – SDI Change from 1870 to 1890



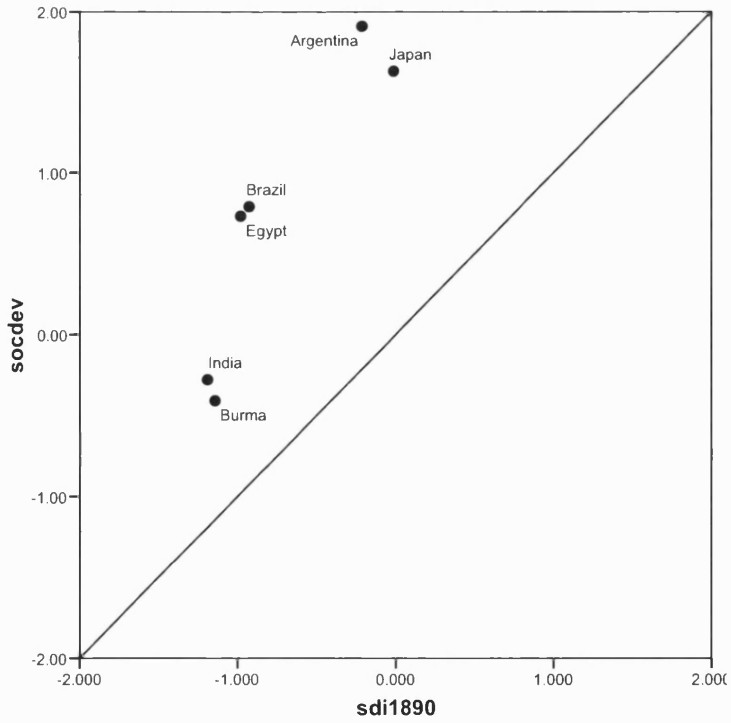
Reference line set equal to the mean =.492

Figure 3.5 - Scatter plot for Social Development Index. Historical Evolution from 1870 to 1960



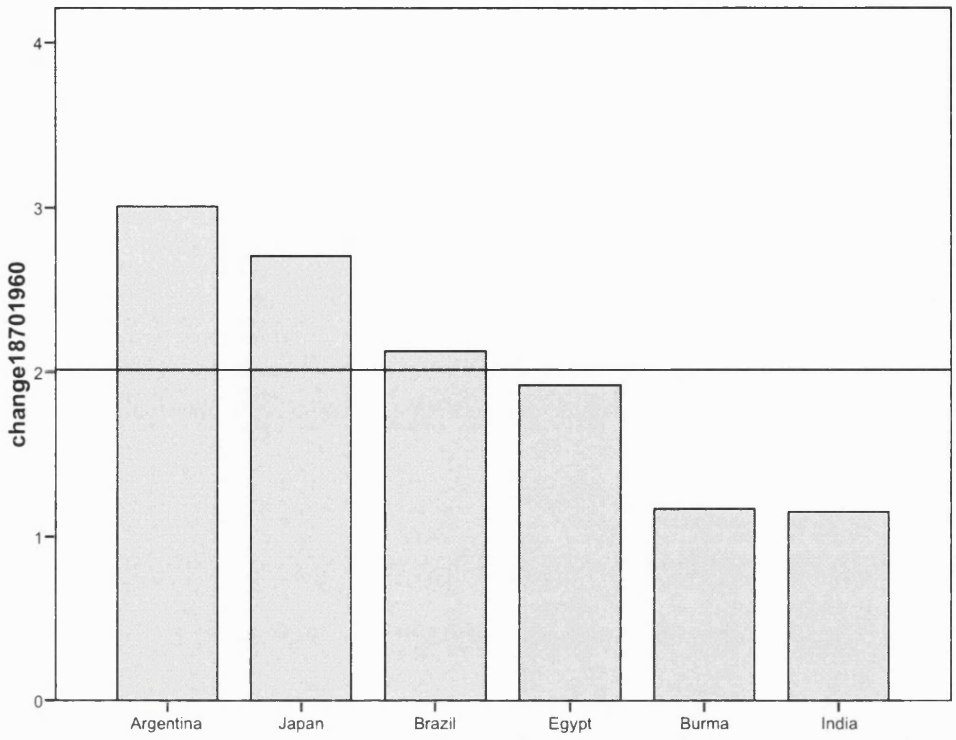
SDI 1870 in the horizontal axis. SOCDEV in the vertical axis.
All countries in the sample improved notably during the period 1870-1960.

Figure 3.6 - Scatter plot for Social Development Index. Historical Evolution from 1890 to 1960



SDI 1890 in the horizontal axis. SOCDEV in the vertical axis.
All countries in the sample improved notably during the period 1890-1960.

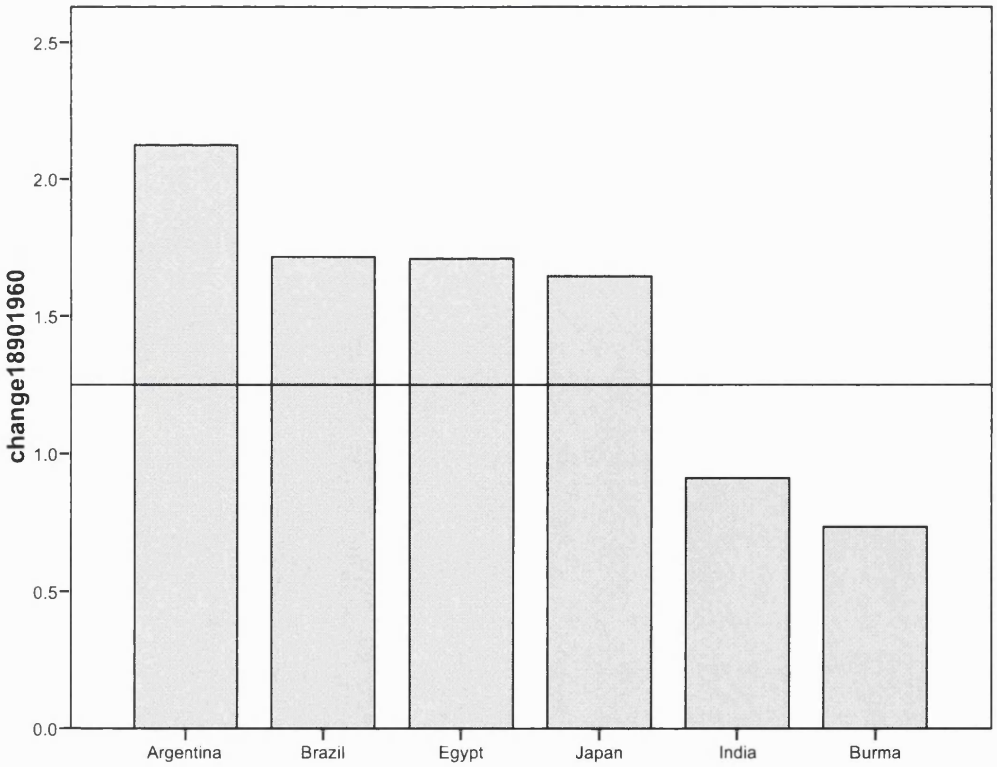
Graph 3.5 - SDI Change from 1870 to 1960



Reference line set at mean value = 2.02

Countries have improved an average of 2 standard deviations between 1870 and 1960 (90 years).

Graph 3.6 - SDI Change from 1890 to 1960

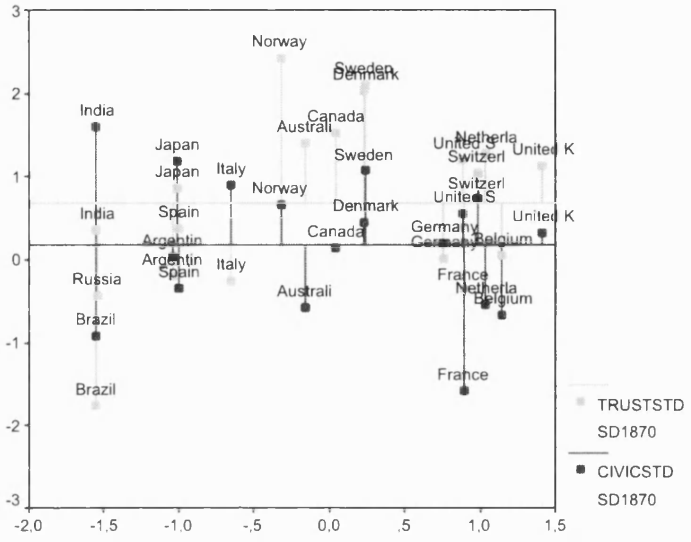


Reference line set at mean value = 1.47

Countries have improved an average of around one and a half standard deviations between 1890 and 1960 (70 years).

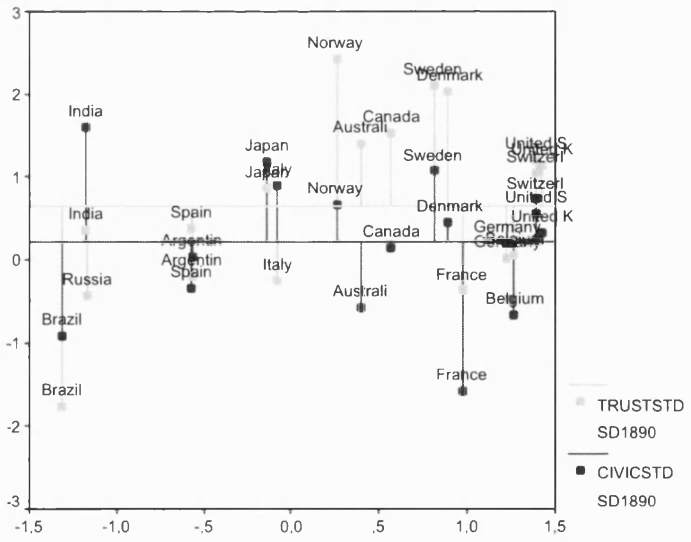
FIGURES FOR SECTION 3.IV - INTERTEMPORAL COMPARISONS: NINETEENTH AND TWENTIETH CENTURIES

Figure 3.7 - Historical Evolution 1870-Nowadays: Overlay Scatter Plot with Spikes



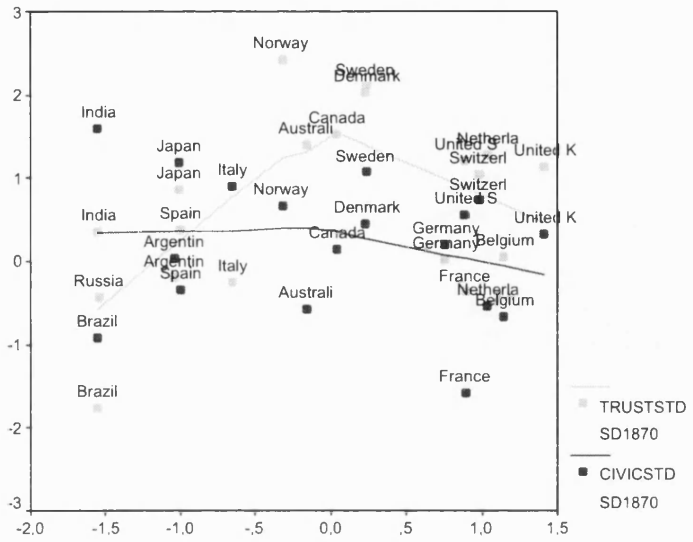
SDI 1870 in the horizontal axis, TRUSTSTD (light) and CIVICSTD (dark) in the vertical axis.

Figure 3.8- Historical Evolution 1890-Nowadays: Overlay Scatter Plot with Spikes



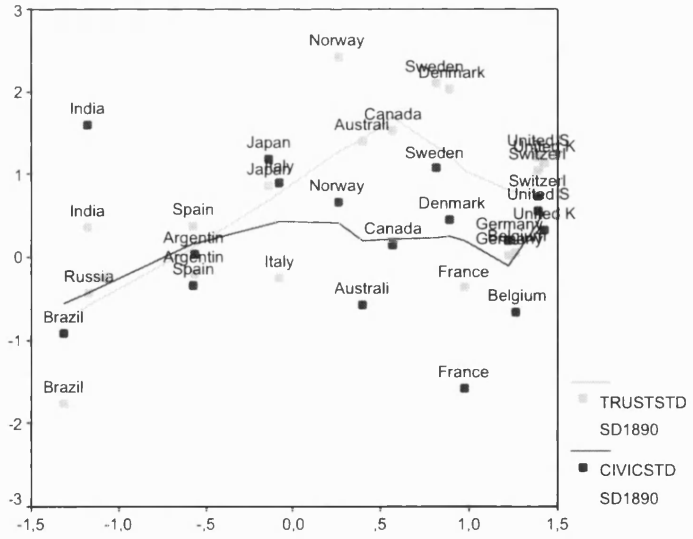
SDI 1890 in the horizontal axis, TRUSTSTD (light) and CIVICSTD (dark) in the vertical axis.

Figure 3.9 - Historical Evolution 1870-Nowadays: Overlay Scatter Plot with Fitted Line



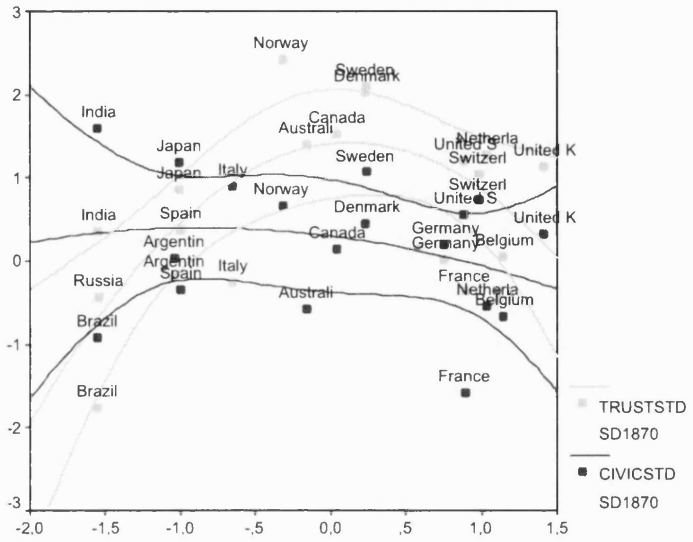
SOCDEV in the horizontal axis, TRUSTSTD (light) and CIVICSTD (dark) in the vertical axis.Fit method: Lowess. 50% of points fitted with 3 iterations.

Figure 3.10 - Historical Evolution 1890-Nowadays: Overlay Scatter Plot with Fitted Line



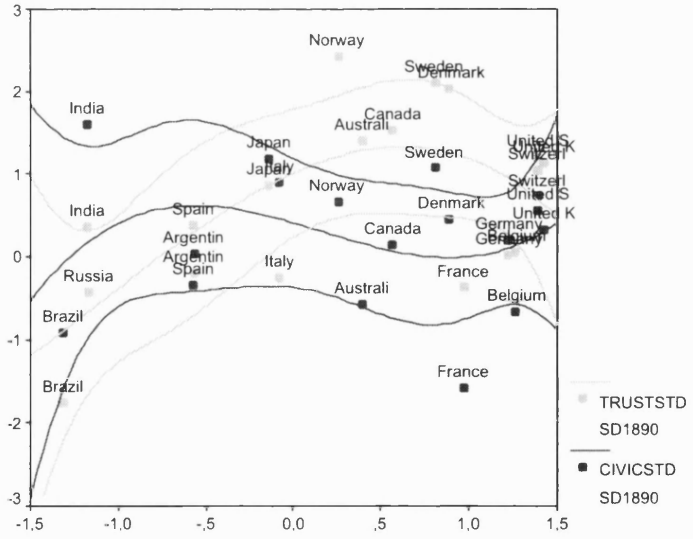
SOCDEV in the horizontal axis, TRUSTSTD (light) and CIVICSTD (dark) in the vertical axis.Fit method: Lowess. 50% of points fitted with 3 iterations.

Figure 3.11 - Historical Evolution 1870-Nowadays. Overlay Scatter Plot with Quadratic Regression Lines



SOCDEV in the horizontal axis, TRUSTSTD (light) and CIVICSTD (dark) in the vertical axis.Fit method: Quadratic regression prediction lines.

Figure 3.12 - Historical Evolution 1890-Nowadays. Overlay Scatter Plot with Cubic Regression Lines



SOCDEV in the horizontal axis, TRUSTSTD (light) and CIVICSTD (dark) in the vertical axis.Fit method: Cubic regression prediction lines.

APPENDIX 3.A

LIST OF COUNTRIES

Recoding of variable 'country' into 'countryn' in order to assign a value to every country:

COUNTRY Old Value	COUNTRYN New Value	Value Label
Argentina	1	Argentina
Australia	2	Australia
Belgium	3	Belgium
Brazil	4	Brazil
Burma	5	Burma
Canada	6	Canada
China	7	China
Denmark	8	Denmark
Egypt	9	Egypt
France	10	France
Germany	11	Germany
India	12	India
Italy	13	Italy
Japan	14	Japan
Netherlands	15	Netherlands
New Zealand	16	New Zealand
Norway	17	Norway
Russia	18	Russia
Spain	19	Spain
Sweden	20	Sweden
Switzerland	21	Switzerland
UK	22	UK
US	23	US

APPENDIX 3.B

KEY TO THE VARIABLES IN THIS CHAPTER

(O) means variable in the adaptation of the original Morris and Adelman (1988) databse. (N) means variable in the new database. (F) means definitive final list of variables that go into SDI. A complete list of variables, including working variables, with sources and construction notes is provided in chapter 2.

O	N	F	Variable Code	Description
x			incgroup	Classification for level of per capita income
	x	x	income	GDP per capita
x			incomgram	Classification for rate of change in per capita income in the past 20 years
	x		incomegr	GDP per capita growth
x	x	x	indutech	Level of development of techniques in industry
x	x	x	indtecgr	Classification for rate of improvement of techniques in industry (lagged, referred to the last 20 years)
x	x	x	agritech	Classification for level of development of techniques in agriculture
x	x		agrtecgr	Classification for rate of improvement of techniques in agriculture
x	x	x	agrilgrp	Classification for percentage of labour force in agriculture
x	x		popxfarm	Population per square kilometres of farmland
x	x	x	intransp	Level of development of inland transportation
x	x	x	transpgr	Classification scheme for rate of improvement of inland transportation (lagged)
x			xgrgroup	Classification scheme for rate of growth of total real exports
	x		trade	Volume of exports relative to GDP
x	x	x	shiftx	Classification for degree of shift in structure of export sector
x	x		indwchan	Classification for direction of change in average real wages in industry
x	x		agrwchan	Classification for direction of change in average real wages or income of the employed agricultural poor
	x		wages	Real wages

x			popgroup	Classification for total population
	x		pop	Total population in thousands
x			popgrgrp	Classification for rate of population growth in the last 20 years
	x		popgr	Cumulative population growth in the last 20 years (%)
x			immigrp	Classification for net immigration
	x		immi	Net migration, in thousands (immigration with positive sign and emigration with negative sign)
	x		mitopopgr	Rate of net migration over total population growth (%)
x	x	x	lit	Classification of extent of adult literacy
x	x		primedgr	Classification for rate of spread of primary education in the past 20 years
x	x		landtenu	Classification for predominant form of land tenure and holding
x	x		landconc	Classification for concentration of landholdings
x	x	x	landadop	Classification for favourableness of land system to adoption of improvements
x	x	x	urbani	Classification for extent of urbanisation
x	x	x	entrep	Classification for favourableness of attitudes towards entrepreneurship
x			govt	Classification for extent of domestic economic role of government in the past 20 years
	x		govrel	Public expenditure as a percentage of GDP
x	x	x	sociopol	Classification for socioeconomic character of national political leadership in the past 20 years
x		x	represen	Classification for strength of national political institutions in the past 20 years
	x		polity2	Revised Polity variable score, ranging from -10 to 10 (from Polity IV, variable under the same name)
x	x		polstabi	Classification for extent of political stability in the past 20 years
x	x		foreignd	Classification for degree of foreign economic dependence in the past 20 years
x	x	x	colstat	Classification for colonial status

x	x	x	mktdev	Component scores for composite indicator of level of development of market institutions up to the given date
x	x	x	mktdevgr	Component scores for composite indicator of rate of spread of market institutions in the last 20 years
x	x	x	mktdvgrl	Component scores for composite indicator of rate of spread of market institutions in the last 20 years (lagged)

APPENDIX 3.C

THE BASICS OF PRINCIPAL COMPONENTS ANALYSIS

The Calculation of the Principal Components

Suppose we have a dataset with n observations on k variables. We denote these variables by $x_1, x_2, \ldots, x_k$. These data can be arranged in a matrix X with n rows and k columns: $X=(x_1\ x_2\ \ldots\ x_k)$.[16] Each column in this matrix contains the data on a particular variable. For instance, the n observations on the first variable, x_1, are in the first column of X. As we have k variables, we can compute all the sample variances ($Var(x_1)$, $Var(x_2)$, ... , $Var(x_k)$) of the variables and all the sample covariances between these variables ($Cov(x_1,x_2)$, $Cov(x_1,x_3)$, etcetera) and arrange them in a variance-covariance-matrix S. This matrix has k rows and k columns. The diagonal of this matrix contains the variances and the other elements of the matrix contain the covariances. For example, the element in the second row and the third column of S contains $Cov(x_2,x_3)$.

The calculation of the Principal components proceeds via the eigenvalues and eigenvectors of the sample covariance matrix S.[17] This matrix can be described by k positive eigenvalues ($\lambda_1, \lambda_2,\ldots,\lambda_k$) and k corresponding eigenvectors ($e_1, e_2,\ldots, e_k$) that have unit length and are orthogonal to each other. Therefore, after calculating the variance-covariance matrix S we can compute (λ_1, e_1), (λ_2, e_2),..., (λ_k, e_k), where we have arranged this sequence in such a way that λ_1 is that largest eigenvalue, λ_2 the one but largest eigenvalue, etcetera. There is a simple relationship between these eigenvalues and eigenvectors of S and the Principal Components.

The relationship is as follows: in the introduction we described the Principal Components as linear combinations of the original variables. A linear combination of the variables $x_1, x_2, \ldots, x_k$ is a weighted sum like

$$P = a_1x_1 + a_2x_2 + \ldots + a_kx_k.$$

[16] In what follows, we assume that the reader is familiar with matrices, and basic matrix multiplication. For more information see Johnson and Wichern (2002), chapter 2.

[17] Eigenvalues and eigenvectors of a matrix are explained briefly in the appendix to this chapter. A more technical discussion can be found in Johnson and Wichern (2002), chapter 2:98-100.

Here, the weights are given by the vector $a=(a_1, a_2, \ldots, a_k)$. Now, it can be shown[18] that the Principal Component that is responsible for the highest variance in the data is the linear combination P_1 with weights equal to the values in the eigenvector of S that corresponds to the largest eigenvalue, i.e. the values in e_1. In short, we obtain the first Principal component by taking the vector a in the formula above equal to the vector e_1. The second Principal Component is the linear combination P_2 with weights equal to the values in the eigenvector of S that corresponds to the one-but-largest eigenvalue, i.e. the values in e_2, and so on and so on. As a result, we will be able to describe the n *observations* we have on the k *variables* of the original data in X as n *factor-scores* on k *Principal Components*. Hence, we effectively transformed an n by k matrix X of data into an n by k matrix $P=(P_1, P_2, \ldots, P_k)$ of factor scores.

At first sight it seems that we have not reduced the data at all. We started with an n by k matrix X and arrive at an n by k matrix P. The difference between X and P, however, is that the columns of P are now independent (orthogonal to each other), each of the columns pointing at a separate independent dimension of variation in the data. Moreover, P is constructed in such a way that its first column (P_1) is the linear combination of the original variables in the data that has the largest variance of all possible linear combinations of the original variables. The second column is the linear combination of the variables in the orginal data that has the second-largest variance, and so on. This fact has important consequences for data-reduction. Indeed, if we remove the last column of P, we know that we remove the direction in the data that has the lowest variance. The resulting columns may well still contain most of the variance that was available in the original data matrix X. Hence, by dropping the last column of P we made sure that at least as possible variation was lost. The same holds for dropping the second-last column and so on. In some cases we can still capture most of the variation in the data by concentrating on only the first or the first two Principal Components.

Geometrical Explanation of Principal Components

Principal Component Analysis is most easily understood graphically in the case where there are only two variables, X1 and X2. Suppose we have a dataset X=(X1 X2) with some observations on these two variables. We can display the variation in these two

[18] See Johnson and Wichern (2002), chapter 8.

variables over different observations in a scatterplot, as in Figure 3.13 below. From this graph we can see that X1 and X2 are positively correlated, and that X1 has a larger variance than X2. If we were interested in reducing the number of variables (data-reduction) while preserving most of the variation in the original data, we would choose to keep X1 and to drop X2. However, we can do better than that.

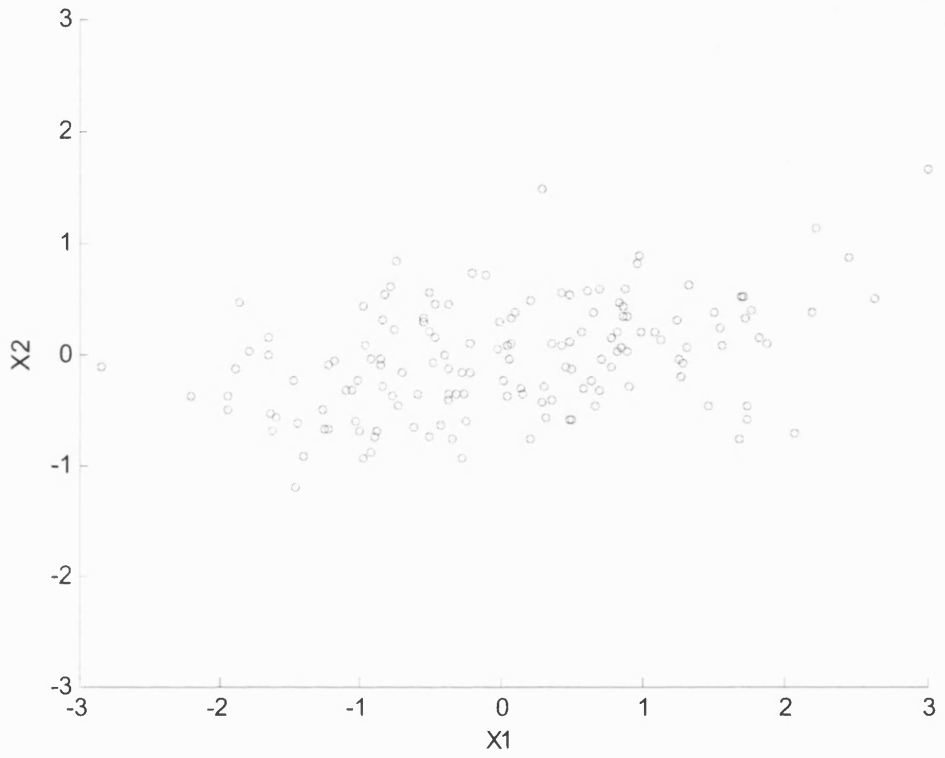


Figure 3.13 - Scatter plot of X1 and X2.

There is positive covariance between X1 and X2, and the variance of X1 is larger than the variance of X2.

In Figure 3.14, two dashed lines are added to the figure. The most horizontal line of the two is the axis along which most of the variation in the data is concentrated. The vector in the graph that follows the direction of this axis turns out to be proportional to the eigenvector e_1 (the eigenvector corresponding to the largest eigenvalue λ_1 of S, the variance-covariance matrix of X). The length of the vector drawn in the graph is exactly λ_1.[19] The second axis that is drawn in the graph runs in the direction of e_2. The length of this vector in the graph is λ_2.

The first axis mentioned above is called the first Principal Component. Note that the points that lie on this axis are simply linear combinations of points on the original axes.

[19] A technical note: the associated contour is the contour that describes the set of points that have statistical distance from the center-point equal to 1.

The second new axis is called the second Principal Component. Note that the two new axes are just a rotated version of the original axes. With respect to these two new axes, all of the points in the graph have new coordinates. Each point now has a

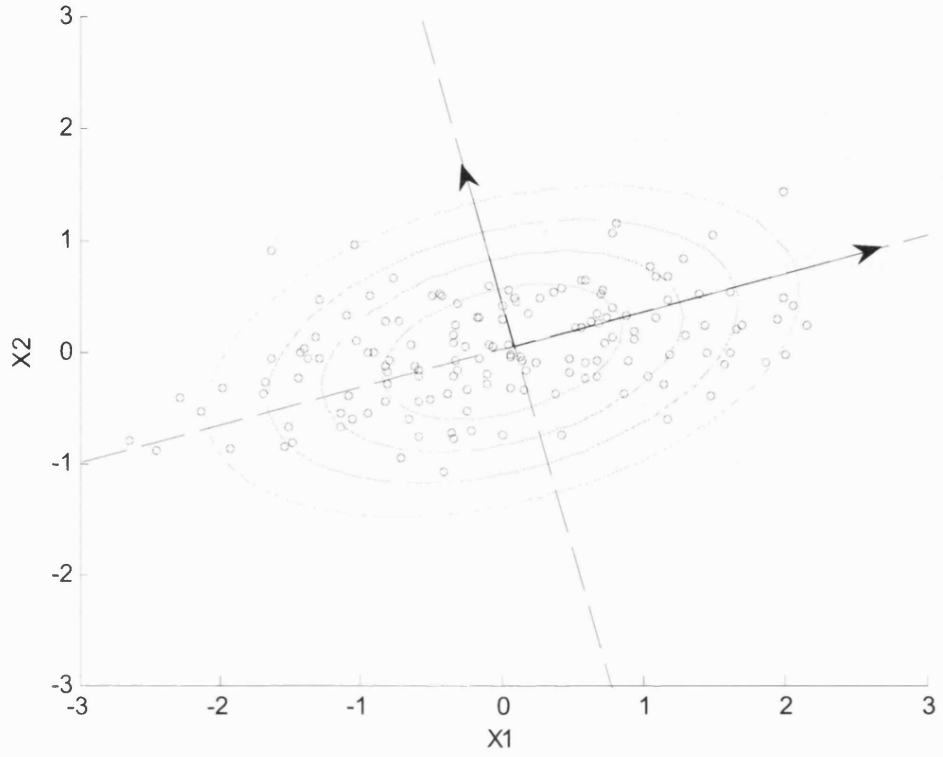


Figure 3.14 - Scatter Plot with Principal Components.
Let S denote the sample variance Covariance matrix of the variables X1 and X2. The eigenvector corresponding to the largest eigenvalue of S is the more horizontal vector in the graph. Its length is the corresponding eigenvalue. Likewise, the more vertical vector is the other eigenvector of S and its length is the smallest eigenvalue of S. Here, the 'most horizontal' axis is the axis with respect to which the data have the widest range of coordinates (factor scores). Therefore, this axis is the first Principal Component. The more vertical axis is the second Principal Component.

coordinate with respect to the first Principal Component and a coordinate with respect to the second Principal Component. These new coordinates are called factor-scores. The idea now is, that if the correlation between X1 and X2 would have been very high, most of the variation in the data would be due to variation in the factor scores on the first Principal Component. Hence, data-reduction would ideally proceed by keeping the first Principal Component and dropping the second Principal Component. This procedure preserves much more variation than the data-reduction strategy mentioned above, where simply X1 was kept and X2 was dropped.

Conceptual Definition of Eigenvalues and Eigenvectors

There are different types of multiplication possible, depending on what is multiplied by what. The simplest case is premultiplication of a variable x by a scalar λ. It is easy to

understand what happens in this case: we obtain λ x, λ times the original variable x. The second type of multiplication we can consider is pre-multiplication of a vector x by a scalar λ. Again, it is easy to understand what will happen to the vector x: we obtain a vector λ x that has the same direction as x, but a length that is λ times the length of the original vector x. The problem occurs in the third type of multiplication that we will consider. It is much harder to understand what happens if we pre-multiply a vector variable x by a matrix A. We obtain the vector Ax, which we will call y.[20] It is typically not clear how the original vector x relates to the resulting vector y. In the abstract, we understand scalar multiplication much better than matrix multiplication.

In order to make the relationship between x and Ax clear, we would really have to calculate the result of multiplying A with x. A priori, we do not have any intuition as to what the result will be. In order to obtain a better intuition of what pre-multiplying with A does to x, it is natural to ask if A maybe acts as scalar multiplication for *some* vectors x. Then, at least for those vectors, we would understand what A does. In other words, we ask ourselves if for a particular matrix A there exists a scalar λ and a corresponding vector x such that $Ax= \lambda x$. If there exists such a scalar and such a vector, we call the scalar λ an eigenvalue of A, and the corresponding vector x the eigenvector of A corresponding to the eigenvalue λ. In particular, to emphasize that this vector x is rather special, we will denote it by e, instead of x. It turns out that for any k by k symmetric matrix A,[21] there actually exist k eigenvalues, $(\lambda_1, \lambda_2, \ldots, \lambda_k)$ and k corresponding eigenvectors $(e_1, e_2, \ldots, e_k)$. These eigenvectors are orthogonal to each other, and can be chosen to have unit length. Details about the calculation of these eigenvalues and eigenvectors can be found in Johnson and Wichern (2002). Most statistical software packages have pre-programmed routines that calculate eigenvalues and eigenvectors for any k by k matrix. The important thing to know about them is that eigenvectors identify the areas (sets of vectors) for which the matrix A works as scalar multiplication. The eigenvalues are the corresponding scalars.

[20] If A is a k by k matrix and x a k by 1 vector, the result y is a k by 1 vector.

[21] Remember that for Principal Component Analysis we are interested in the eigenvalues and eigenvectors of the sample covariance matrix S. This matrix is symmetric.

APPENDIX 3.D

PRE-PCA PRELIMINARY TESTS:

TABLES FOR SECOND BEST ALTERNATIVES

Table 3.14 – Kaiser-Meyer-Olkin (KMO) sample adequacy scores variable-by-variable, Second Best alternatives

Variables	Original Data Criterion: KMO≥.5		All New Data* Criterion: KMO≥.6
	2nd Round	3rd (and final) round	2nd (and final) round
Income	.7140	.6883	.8544
income growth	.6397	.6341	-
Technique in industry	.7663	.7928	.9074
adoption of new techniques	.8745	.8907	.8931
Technique in agriculture	.8128	.7966	.9359
agricultural labour	.7574	.7511	.8292
population per farmland	.4896	-	-
inland transportation	.8317	.8909	.9378
transportation, growth	.7403	.8078	-
export growth group	-	-	-
Trade	-	-	-
shift in export structure	.7833	.8362	.9079
industrial wage change	.6503	.7454	-
agricultural wage change	.6588	.6674	-
Wages	-	-	-
population	.4850	-	-
population growth	-	-	-
immigration	-	-	-
Migration to pop. growth	-	-	-
Literacy	.8056	.7354	.8690
primary education growth	.6565	.6059	-
form of land tenure	.6260	.5573	-
land concentration	-	-	-
land adoption	.7227	.7141	.9240
urbanisation	.8963	.8715	.9240

entrepreneurship	.8856	.8808	.9471
role of government	.5757	.6140	-
socio-politics	.8768	.8778	.9370
representativeness	.8176	.7891	-
polity 2	-	-	-
political stability	.6482	.6354	.7800
foreign dependency	.8160	.8164	.8479
colonial status	.6930	.6820	.7741
market development	.8347	.8448	.9060
market development growth	.8037	.7882	.8588
mkt develpt growth, lagged	.7728	.7570	.8551
Number of variables	**28**	**26**	**18**
Overall sampling adequacy	**.7661**	**.7694**	**.8859**

KMO measures sampling adequacy, in a scale from 0 to 1.
* All new data refers to the new database keeping goverel and Polity2, which have been replaced by the old variables due to better performance in the first best option.

Table 3.15 - Squared Multiple Correlations (SMC) of variables with all other variables Second Best alternatives

Variables	Original Data Criterion: KMO≥.5		All New Data* Criterion: KMO≥.6
	2nd Round	3rd (and final) round	2nd (and final) round
Income	.9796	.9793	.9341
income growth	.8749	.8503	-
Technique in industry	.9650	.9478	.9180
adoption of new techniques	.9189	.9055	.8592
Technique in agriculture	.9586	.9548	.9048
agricultural labour	.9432	.9411	.9038
population per farmland	.8279	-	-
inland transportation	.9395	.9072	.8610
transportation, growth	.7539	.7135	-
export growth group	-	-	-
Trade	-	-	-
shift in export structure	.8415	.8038	.6853
industrial wage change	.7833	.7163	-
agricultural wage change	.6009	.5864	-
Wages	-	-	-
population	.8841	-	-
population growth	-	-	-
immigration	-	-	-
Migration to pop. growth	-	-	-
Literacy	.9736	.9684	.9165
primary education growth	.7783	.7756	-
form of land tenure	.9000	.8735	-
land concentration	-	-	-
land adoption	.9235	.9122	.7664
urbanisation	.8596	.8583	.7749
entrepreneurship	.9546	.9469	.8979
role of government	.6916	.6552	-
socio-politics	.9021	.8912	.8536
representativeness	.9390	.9368	-

polity 2	-	-	-
political stability	.9407	.9264	.8285
foreign dependency	.8850	.8590	.7502
colonial status	.9274	.9208	.8265
market development	.9692	.9657	.9394
market development growth	.9982	.9981	.9973
mkt develpt growth, lagged	.9982	.9981	.9970
Number of variables	**28**	**26**	**18**
Overall sampling adequacy	**.7661**	**.7694**	**.8859**

* All new data refers to the new database keeping goverel and Polity2, which have been replaced by the old variables due to better performance in the first best option.

REFERENCES

Abramovitz, Moses, 1986, 'Catching Up, Forgoing Ahead, and Falling Behind', *The Journal of Economic History*, 46 (2):385-406.

Acemoglu, Daron, Simon Johnson, and James A. Robinson, 2001, 'The Colonial Origins of Comparative Development: an Empirical Investigation', *American Economic Review*, 91 (5): 1369-1401.

Adelman, Irma, and Cynthia T. Morris, 1965, 'A Factor Analysis of the Interrelationship Between Social and Political Variables and Per Capita Gross National Product', *The Quarterly Journal of Economics*, 79 (4): 555-578.

———, 1967, *Society, Politics, and Economic Development: A Quantitative Approach*, John Hopkins Press (Baltimore).

———, 1978, 'Patterns of Market Expansion in the Nineteenth Century: A Quantitative Study', in George Dalton, ed., *Research in Economic Anthropology*, JAI Press (Greenwich, Conn.): 231-324.

Alesina, Alberto and Eliana La Ferrara, 2000, 'Who Trusts Others?', CEPR Discussion Paper Series, 2646.

Allen, Robert, 2004, 'Progres and Poverty in Early Modern Europe', Seminar series at the IHR (London).

Anderson, Eugene Newton, and Pauline Safford R. Anderson, 1967, *Political Institutions and Social Change in Continental Europe in the Nineteenth Century*, University of California Press (Berkeley and Los Angeles).

Anderson, Perry, 1974, *Lineages of the Absolutist State*, New Left Books (London).

Barro, Robert J. and Jong-Wha Lee, 'International Measures of Schooling Years and Schooling Quality', *American Economic Review*, Papers and Proceedings 86 (2), 1996, pp. 218-23. (Updated dataset at: http://www.worldbank.org/research/growth/ddbarle2.htm)

Bartlett, 1950, 'Test of Significance in Factor Analysis', *British Journal of Psychology*, 3:77-85.

Blyth, Mark, 2002, *Great Transformations: Economic Ideas and Institutional Change in the Twentieth Century*, Cambridge University Press (Cambridge,UK).

Cairncross, Alexander Kirkland, 1962, *Factors in Economic Development*, Allen & Unwin (London).

Cohen, Daniel and Marcelo Soto, 'Growth and Human Capital: Good Data, Good Results', CEPR discussion paper 3025, (2001).

Davis, J. A. and T. W. Smith, 1994, *General Social Surverys, 1972-1994*, produced by the National Opinion Research Center (Chicago), distributed by The Roper Center for Public Opinion Research (Storrs, Connecticut).

De la Fuente, Angel and Rafael Doménech, 'Human Capital in Growth Regression: How Much Difference Does Quality Data Make', CEPR discussion paper 2466, (2000).

Deutch, Karl W., 1953, *Nationalism and Social Communication: An Inquiry Into the Foundations of Nationality*, MIT Technology Press and John Wiley & Sons (Cambridge, MA).

Diana, Giancarlo, and Chiara Tommasi, 2002, 'Cross-validation Methods in Principal Components Analysis: A Comparison", *Statistical Methods & Applications*, 11:71-82.

DiPasquale, Denise, and Edward L. Glaeser, 1999, 'Incentives and Social Capital: Are Homeowners Better Citizens?' *Journal of Urban Economics*, 45: 354-384.

Furnival, John S., 1931, *An Introduction to the Political Economy of Burma*, Burma Book Club (Rangoon).

Gerschenkron, Alexander, 1962, *Economic Backwardness in Historical Perspective: A Book of Essays*, Harvard University Press (Cambridge, MA).

Groningen Growth and Development Centre (GGDC) and The Conference Board, *Total Economy Database*, January 2005, http://www.ggdc.net

Hanley, Susan Bell, and Yamamura, Kozo, 1977, *Economic and Demographic Change in Preindustrial Japan, 1600-1868*, Princeton University Press (Princeton).

Harriss, John, 2002, *Depoliticizing Development: The World Bank and Social Capital*, Anthem Press (London).

Helleiner, Eric, 1994, *States and the Reemergence of Global Finance*, Cornell University Press (Ithaca).

Heston, Alan, Robert Summers and Bettina Aten, *Penn World Table Version 6.1*, Center for International Comparisons at the University of Pennsylvania (CICUP), October 2002.

Johnson, Richard A. and Dean W. Wichern, 2002, *Applied Multivariate Statistical Analysis*, Prentice-Hall (London).

Kaiser, Henry F., 1970, 'A Second Generation Little Jiffy', *Psychometrika*, 35 (4, December):401-415.

Kirshner, Jonathan, 1999, 'Keynes, Capital Mobility and the Crisis of Embedded Liberalism', *Review of International Political Economy* 6 (3, Autumn):313-337.

Knack, Stephen, and Philip Keefer, 1997, 'Does Social Capital Have an Economic Payoff? A Cross-Country Investigation', *The Quarterly Journal of Economics*, CXII:1251-1288.

Krueger, Alan B. and Mikael Lindahl, 'Education for Growth: Why and For Whom', NBER working paper 7591 (2000).

La Porta, Rafael, Florencio Lopez-de-Silanes, Andrei Shleifer, and Robert W. Vishny, 1997, 'Trust in Large Organisations', *American Economic Review Papers and Proceedings*, 87: 333-338.

Lee, William Robert, 1977, *Population Growth, Economic Development, and Social Change in Bavaria, 1750-1859*, Arno Press (New York).

Lindert, Peter, 2003, "Voice and Growth: Was Churchill Right?" *Journal of Economic History* 63, 2 (June): 315-350.

MacNamara, Kathleen R., 1998, *The Currency of Ideas: Monetary Politics and the European Union*, Cornell University Press (Ithaca): esp. 54-5, 82-7.

Maddison, Angus, 2003, *The World Economy: Historical Statistics*, OECD Development Centre (Paris).

Manli, Bryan F. J., 1991, *Multivariate Statistical Methods: A Primer*, Chapman and Hall (London).

Milward, Alan S., and S. B. Saul, 1979, The *Economic Development of Continental Europe, 1780-1870*, 2nd ed, George Allen & Unwin, (London).

Moore, Barrington, Jr., 1966, *Social Origins of Dictatorship and Democracy: Lord and Peasant in the Making of the Modern World*, Beacon Press (Boston).

Morris, Cynthia Taft and Irma Adelman, 1988, *Comparative Patterns of Economic Development, 180-1914*, The John Hopkins University Press (Baltimore and London).

Mukerjee, Tapan, 1972, 'Theory of Economic Drain: Impact of British Rule on the Indian Economy, 1840-1900' in *Economic Imperialism: A Book of Readings.*

North, Douglass, 1981, *Structure and Change in Economic History*, Norton, New York.

North, Douglass Cecil, 1990, *Institutions, Institutional Change and Economic Performance*, Cambridge University Press (Cambridge, UK).

Polanyi, Karl, 1944, *The Great Transformation: The Policial and Economic Origins of Our Time*, Farrar & Rinehart (New York), reprinted in 1957 by Beacon (Boston).

Pollard, Sidney, 1988, Review of Cynthia Taft Morris and Irma Adelman, 1988, *Comparative Patterns of Economic Development, 180-1914*, The John

Hopkins University Press (Baltimore and London), *The Economic History Review* 41 (4):668-670.

Putnam, Robert D., *Making Democracy Work: Civic Traditions in Modern Italy*, (Princeton University Press, 1993).

—— , *Bowling Alone: The Collapse and Revival of American Community*, Simon & Schuster, (New York, 2000).

Read, Ronald C., 1972, *A Mathematical Background for Economists and Social Scientists*, Prentice-Hall (Englewood Cliffs, NJ).

Ruggie, John Gerald, 1982, 'International Regimes, Transactions, and Change: Embedded Liberalism in the Postwar Economic Order', *International Organization* 36 (2, Spring):379-415.

Simon, Carl P. and Blume, Lawrence, 1994, *Mathematics for Economists*, W. W. Norton & Company (New York and London).

STATA User's Manual.

Tabellini, Guido, 2007, 'Culture and Institutions: Economic Development in the regions of Europe', http://www.igier.uni-bocconi.it/whos.php?vedi=327&tbn=albero&id_doc=177 .

Temple, Jonathan, and Paul A. Johnson, 'Social Capability and Economic Growth', *The Quarterly Journal of Economics* 113 (3), 1998, pp. 965-990.

Warren, Bill, 1980, *Imperialism: Pioneer of Capitalism*, New Left Books (London).

Wuensch, Karl L., 2005, Principal Components Analysis – SPSS, East Carolina University, Department of Psychology, http://core.ecu.edu/psyc/wuenschk/MV/FA/PCA-SPSS.doc

Chapter 4

THE SOCIAL DEVELOPMENT INDEX AND PER CAPITA INCOME: TESTING FOR A STRUCTURAL RELATIONSHIP IN THE LONG RUN

4. I. INTRODUCTION

Social capital is a form of capital. As such, it has been shown in previous studies that it is bonded to economic performance. To mention some of the most influential papers, La Porta *et al.* (1997), Knack and Keefer (1997), and Temple and Johnson (1998) all tested the statistical significance of social capital indicators in explaining economic performance. For the relatively recent periods they analysed, they all found a positive association between the two.

First, La Porta *et al.* find supporting evidence for the fact that trust and cooperation have a positive effect on the well-functioning of organisations such as government, professional societies, and large firms. More generally, they show that social capital makes institutions work better; at least for the period 1970-1993.

Second, Knack and Keefer (1997) test the significance of social capital in explaining economic growth for the period 1980-1992. For this, they use two alternative measures, trust and the civic engagement index, both extracted from the World Value Surveys. Growth regressions incorporating these variables show a strong effect of both of them on economic growth. A 10 percent increase in the level of trust is associated with a 0.8 percent increase in 1980-1992 growth. A similar increase in the civic engagement index corresponds to a 2.7 percent increase in economic growth (Knack and Keefer, 1997:1260-1).

Finally, Temple and Johnson (1998) use the dataset and sample of Mankiw, Romer and Weil (1992) for measures of economic prosperity. In particular, they contrast the 1960's social development index constructed by Adelman and Morris (1967) for another purpose with the log difference of per capita income 1960-1985 (long run growth of GDP). They found a correlation coefficient of 0.60 (Temple and Johnson, 1998:970). When they run ordinary least squares (OLS) growth regressions, the social development index coefficient is statistically significant at the 0.05 significance level. This result is robust to the inclusion of initial income, investment, schooling, and population growth in the growth regressions. This finding supports the initial hypothesis of the predictive power of social development in explaining subsequent economic growth.

The Adelman and Morris (1967) work could shed some light on the questions in the agenda. Now that we can look at their analysis in historical perspective, it appears to have not only the historical insights the authors were presenting but also more predictive power than they suspected. When they created them, their indicators turned out to be *apparently* too pessimistic for Latin America and too optimistic for some Far Eastern countries. At that time, they did not know the importance of their findings. Their social indicator could have helped forecasting subsequent growth better than any other contemporary attempt. Temple and Johnson (1998) tell us that the socio-economic index constructed by Adelman and Morris 'could have helped researchers make much better forecasts of long-run growth rates'.

Notice that all of these studies have a fairly recent chronology. The earliest one is the study by Temple and Johnson (1998), which starts in 1960. This is the only one using the Adelman and Morris socio-development indicator; therefore, it provides us with a framework for comparison.

The nineteenth century social capital index proposed in Chapter 3 is based on own calculations along the lines of Adelman and Morris. The variables are not exactly the same due to historical data limitations, but they have been constructed keeping their spirit in mind. We make use of another well-established international database for socio-economic measures from 1850 to 1913, broken into three periods: 1850-1870, 1870-1890, and 1890-1913. It comprises a wide range of socio-economic variables, 35 in total, ranging from education or political attitudes to favourableness towards adoption of new technology. Morris and Adelman (1988) worked for more than twenty years (since 1965) on this database together with a team of country experts so that the data were homogeneous and reliable. The data needed to be recovered, digitalised, and updated. Although the original data were not gathered for this purpose, we propose using this international dataset for the nineteenth century to construct a completely new Social Development Index (SDI), based on a principal components analysis, and described in detail in Chapter 3.

The Social Development Index (SDI) can be interpreted as measuring social development or social capital, just as the Human Development Index (HDI) measures human development. The new SDI series are the first available nineteenth-century international social capital estimates. They are available for a large set of countries, 22

in total, from all continents. We make use of the SDI in this paper in order to investigate its economic implications.

The new Social Development Index (SDI) gives us the possibility to do the tests that Temple and Johnson performed on 1960 data, but now on data from the nineteenth century; in particular, we do it for 1870 and 1890. By doing so, we could determine whether the relationship between social development and economic performance could have further theoretical implications beyond the post-Second World War era. The question to answer in this chapter is the following: Is the relationship between social capital and economic performance consistently present over time? In other words, is there some evidence for a long run structural relationship between social capital and macroeconomic performance.

4. II. THE RELATIONSHIP BETWEEN SDI AND INCOME

In order to answer the question of whether the relationship between social capital and economic performance is consistent over time, let us start with a snapshot of the work by Temple and Johnson mentioned in the previous section. This work refers to the period 1960-1985.

Figure I and II have been extracted from Temple and Johnson (1998:971-2). In both figures, the social development index derived in Adelman and Morris (1967) for the period 1957-62 is displayed on the horizontal axis. In short, their indicator is a compound index extracted from a selection and combination of variables belonging to a broad database with 41 variables and 72 countries (see Adelman and Morris, 1967 for details). In Figure I, this index is plotted against the natural logarithm of GDP in 1960, the initial GDP for the period under study. The figure shows a clearly positive relation between initial GDP and social development. The second figure plots social development against long run per capita GDP growth rate and again we see a positive association between the two. Therefore, according to the work of Temple and

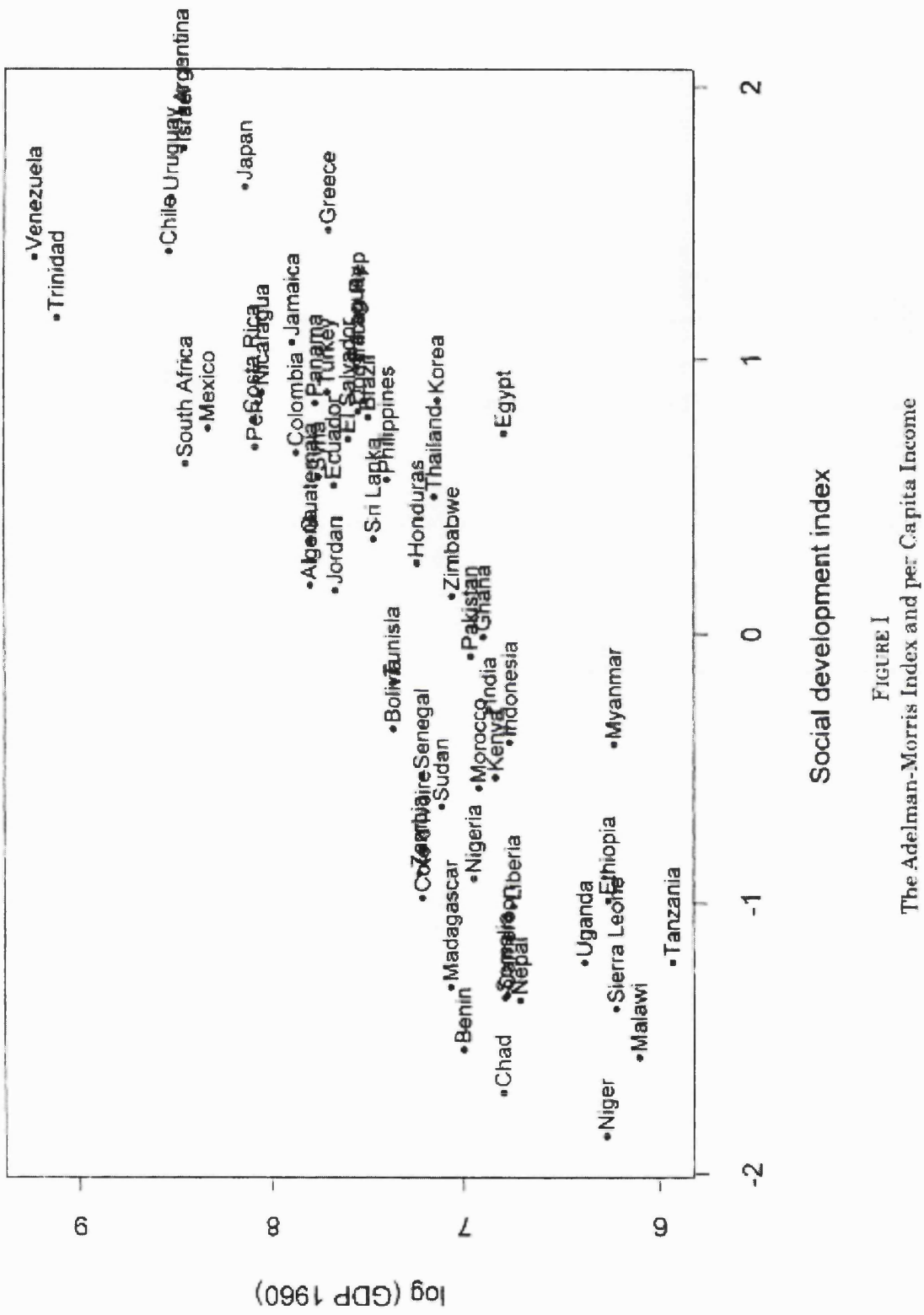


FIGURE I
The Adelman-Morris Index and per Capita Income

Source: Temple and Johnson (1998:971)

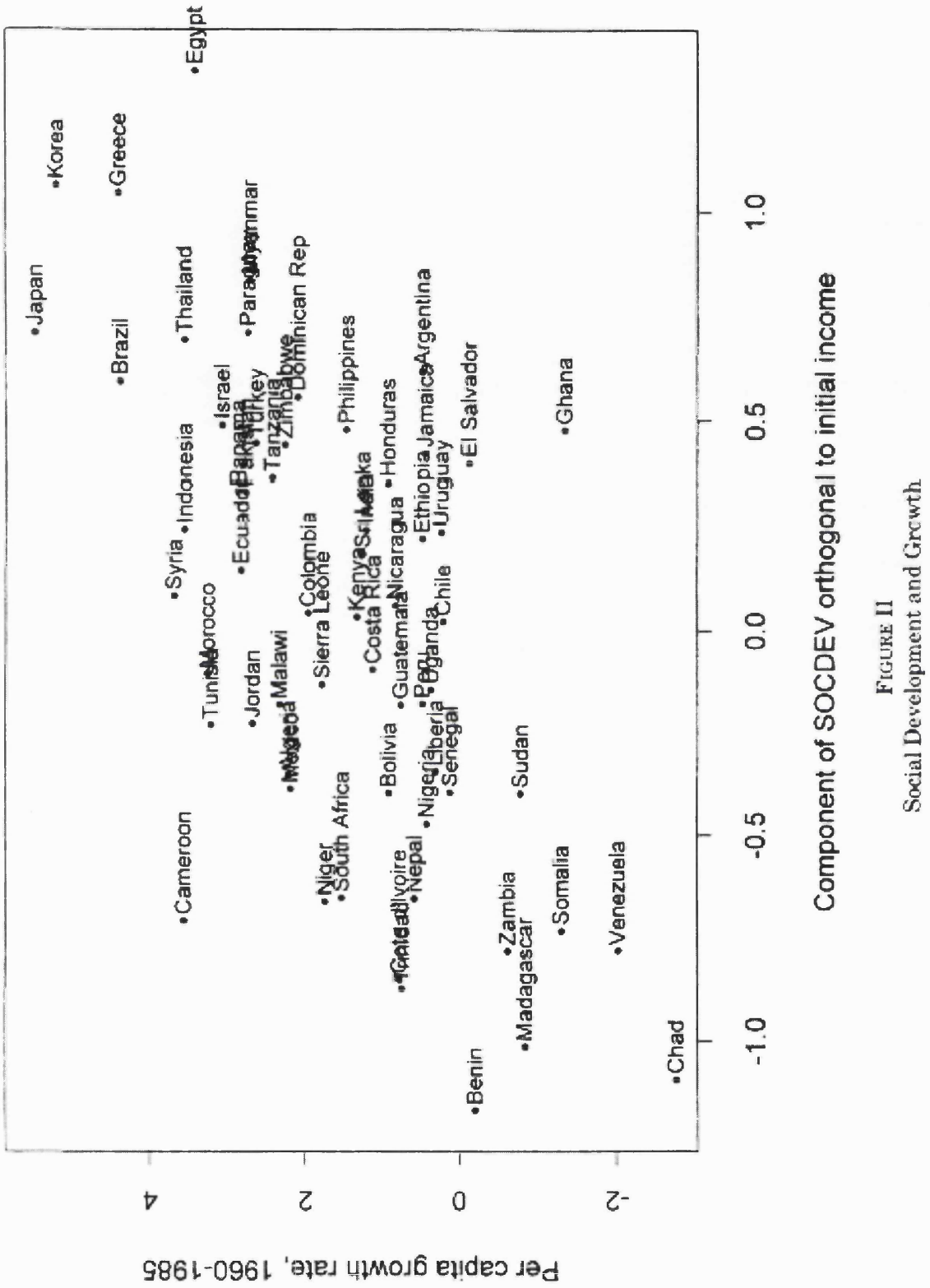


FIGURE II
Social Development and Growth

Source: Temple and Johnson (1998:972)

Johnson, higher values of social development are associated with higher values of both initial GDP and subsequent GDP per capita growth for the period 1960-85.

As described in the previous section, we now have nineteenth century data available, in particular for the years 1870 and 1890. Chapter 3 presented a social development index for 1870 and 1890. The natural question that appears, then, is whether the associations between social development and GDP can be found as well for the nineteenth century. More generally speaking, one would like to compare the results obtained for the nineteenth century to those available for the twentieth century in order to identify any relevant changes in the relationship.

4. II. 1. THE TEMPLE AND JOHNSON TESTS

In this section, we first replicate Figure I of Temple and Johnson for 1960 and, second, we test for a similar relationship between social development and GDP per capita in the nineteenth century.

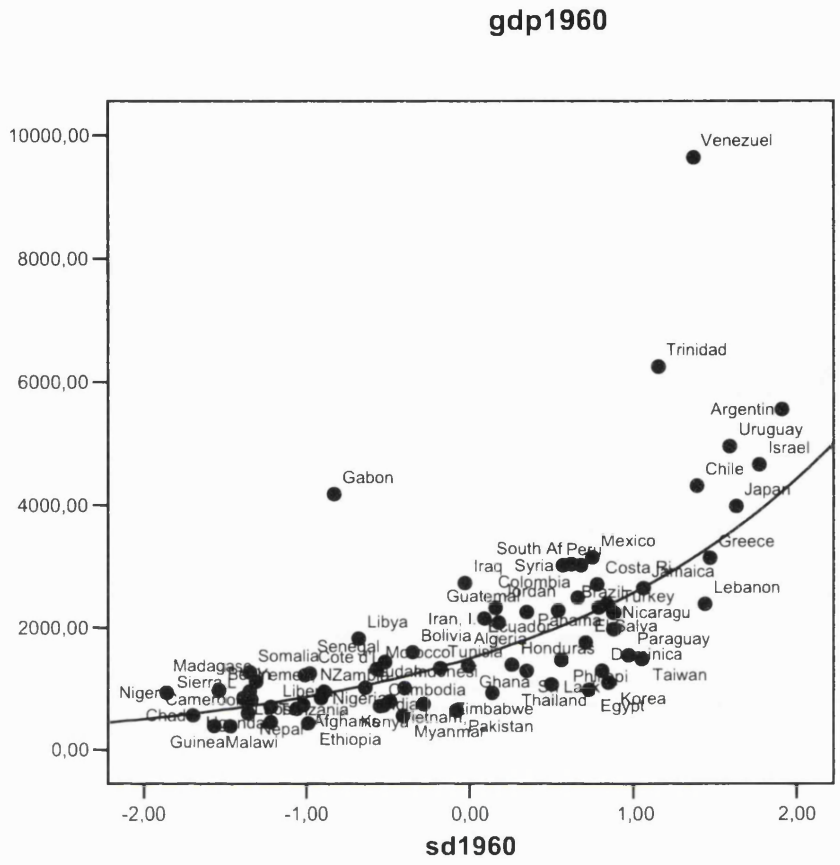
In order to allow for comparison, we need to contrast data for 1870, 1890 and 1960 (the 3 years for which we have SDI) using a comparable GDP data source. This is to avoid that changes in the results between the two centuries can be attributed to inconsistencies in sourcing. Temple and Johnson (1998) use the Summers and Heston database for GDP and other economic measures, which starts in 1950. More precisely, instead of the original they use a secondary source, Mankiw, Romer, and Weil (1992), which starts at 1960, presumably to obtain full data for a larger number of countries. Therefore, if we are to replicate the Temple and Johnson tests for the twentieth century and present new ones for the twentieth century with a consistent data source for GDP, we cannot use the Summers and Heston data. For this reason, the latest 2007 updated version of Maddison (2003) historical statistics on per capita GDP has been used instead. The unit of measure is per capita GDP, 1990 International Geary-Khamis dollars. There is an ongoing discussion about the conveniences and inconveniences of using this way of adjusting GDP, but this source has the advantage that it contains all periods and countries in the sample.[1]

[1] For a discussion on the advantages and disadvantages of using Maddison's data for GDP, see Chapter 2.

Graph 4.1 plots social development for 1960 against the raw level of GDP per capita in 1960. This graph seems to suggest an exponential relationship between SDI and income, which is expected: An exponential relationship between social development and income implies a linear relationship between social development and the natural logarithm of income, as seen in Figure I of Temple and Johnson (1998). An exponential estimation line has been added to the graph and, indeed, fits the data well (R-squared equals 0.611). As a consequence, we should find that the relationship between the index and the log of GDP per capita is linear. We now investigate this relationship more rigorously.

Graph 4.2 is a replica of Temple and Johnson's Figure I using 2007 update of Maddison (2003) data for GDP.[2] There are 74 country observations in this scatter plot, those corresponding to the 74 countries in the Adelman and Morris (1967) sample. Although the exact figures for Temple and Johnson were not reported, it is worth noting that the scatter plot is nearly identical to the scatter plot of Temple and Johnson. So we can conclude that the change of data source and reference year has no important effect on the results. A linear trend fitted through the cloud of points returns a highly statistically significant slope of 0.549 (see Table 4.1, equation 6). This means that an increase of one whole unit in the 1960 social development index (ranging approximately between -2 and 2) is associated with a log GDP level 0.549 units higher. Another way of seeing moves along the line is measuring moves in terms of standard deviations: The standardised beta coefficient (not shown in the Table but equal to 0.781) implies that one standard deviation increase in SDI corresponds to a 78.1 percent of a standard deviation increase in log GDP.

[2] Maddison (2005) has been the source for GDP figures of Cyprus and Suriname, which do not appear in his later work. GDP for Cyprus 1960 and 1985 has been interpolated from the existing 1950 and 1990 estimations at a constant rate of growth; likewise for Suriname 1960.

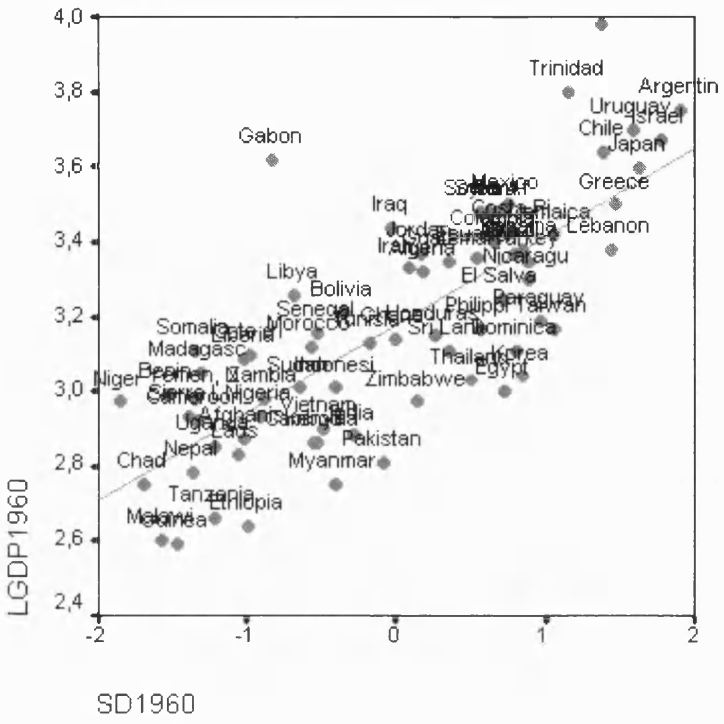


Graph 4.1 – Social Development and Per Capita GDP in 1960

This graph illustrates the exponential relation between SDI and per capita income. An exponential line with a constant has been fitted to the scatter plot, obtaining an R-squared of 0.611.

The adjusted R-squared is 0.605; i.e. for this sample of countries, 60.5 percent of the total international dispersion of incomes in 1960 can be captured with only one explanatory variable (social development) and a constant. Caution: this does *not* necessarily imply that social development *causes* 60.5 percent of the variation of income. This figure is just statistical correlation; we should introduce some control variables in a multivariate regression to find out the clean effect of one single variable on another; (this is done later on in the chapter).

We now move to the nineteenth century analysis, making use of the new SDI series for 1870 and 1890. We have found that in the post-Second World War era higher



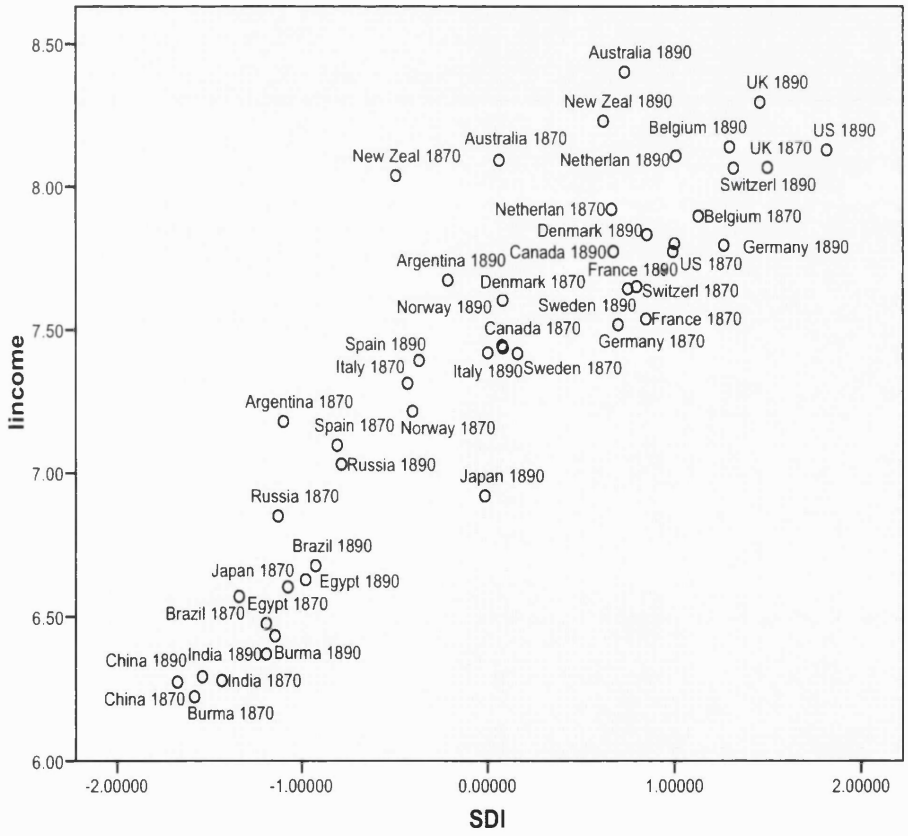
Graph 4.2 – Social Development and Log Per Capita GDP in 1960

levels of social development are associated with higher levels of income and vice versa. Given that we have a social development index for 1870 and 1890 constructed earlier in the thesis, we can now investigate whether this relationship did already exist in the late-nineteenth century. Graph 4.3 shows that a very similar relationship is found for the late nineteenth century, for which we have data. Graphs 4.4 and 4.5 illustrate this relationship separately for years 1870 and 1890 respectively. Table 4.1 corroborates the results statistically, and shows that the figures obtained for 1870 and 1890 not only go in the same direction but are also very similar to those obtained for 1960.

Regression (1) in Table 4.1 is a pooled sample of the 1870 and 1890 observations considered together in order to increase the sample size. This alternative is not unreasonable, since the 2 time cuts for the nineteenth century are not too far apart; so within the framework of a long run perspective they could –if necessary– be considered as belonging to the same *époque*. The result is a sample of 46 observations giving an average SDI coefficient of 0.575 (highly significant) and almost 80 of the total variance explained with the social development index only. There is no need to include time fixed effects in this panel regression, since the differential effects of the 2 time cuts have already been estimated separately in regressions (2) to (5) and the only purpose of this

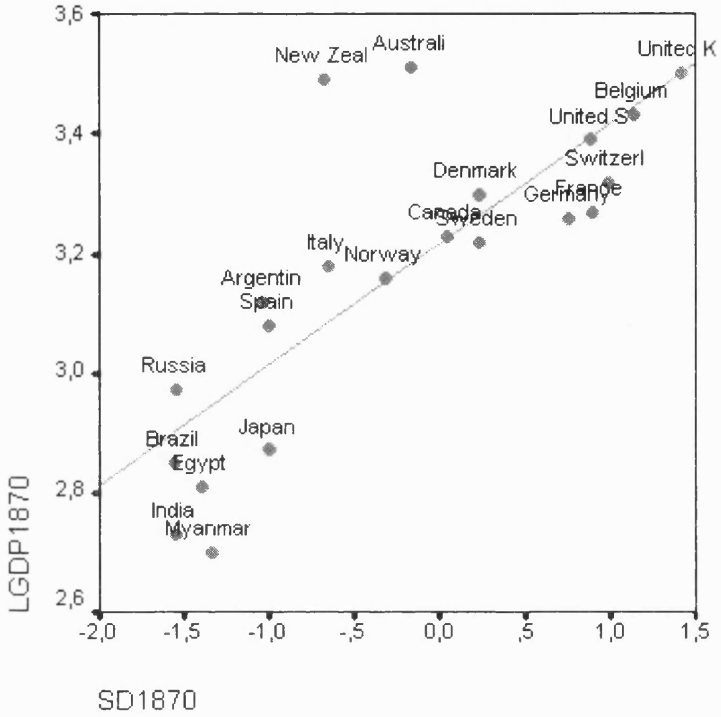
regression was to test for robustness of the results in a larger sample. Graph 4.3 is the scatter plot corresponding to the pooled nineteenth century sample.

Nineteenth century calculations for individual years count only with 23 countries due to narrower data availability. Still, statistical results for bivariate analysis are surprisingly clear cut and they all point at the same direction. The line fitted to Graph 4.4 for 1870 exhibits a 54 percent slope; the line fitted to Graph 4.5 for a 1890 parallel analysis exhibits a slope of 62 percent. Recall that the slope for 1960 data was just below 55 percent. All three analyses have highly statistically significant and very similar correlation coefficients. Moreover, all display fairly high adjusted R-squares varying between 0.61 and 0.84; i.e., the linear regressions fitted on the scatter plots can explain between 61 and 84 percent of dispersion.

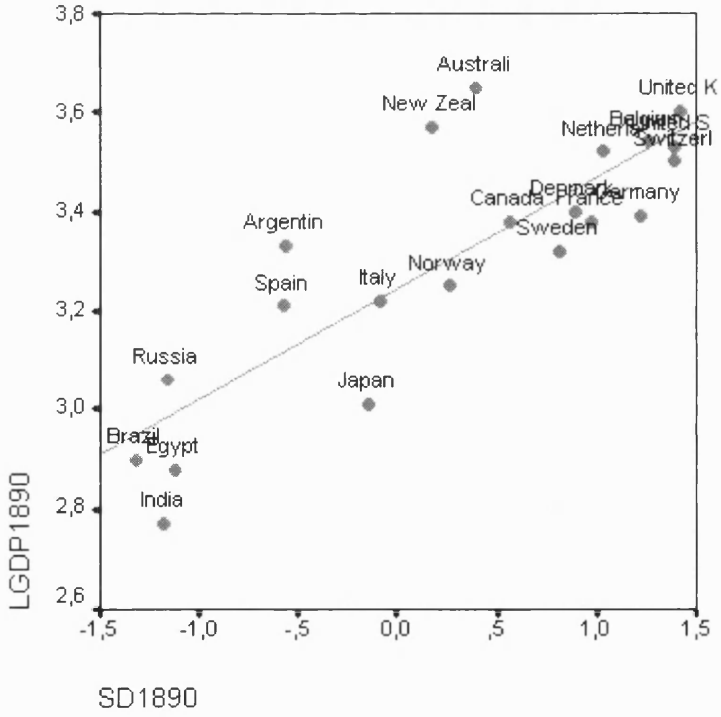


Graph 4.3 –

Social Development and Log Per Capita GDP in the Late-Nineteenth Century, Pooled Sample



Graph 4.4 – Social Development and Log Per Capita GDP in 1870



Graph 4.5 – Social Development and Log Per Capita GDP in 1890

Adjusted R-squares raise up to more than 88 percent of dispersion captured when we remove 1-2 influential observations from each of the regressions (see equations 3, 5, and 7 in Table 4.1). Still, slopes remain practically unchanged in all three chronologies.

Table 4.1
Level of association between Social Development and Income
Dependent Variable: Log GDP Method: OLS

Regression	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Year	Pooled 1870& 1890	1870	1870	1890	1890	1960	1960
Observations	46	23	21	23	22	74	72
SDI slope	.575*** (.044)	.543*** (.070)	.544*** (.044)	.620*** (.059)	.605*** (.053)	.549*** (.052)	.544*** (.046)
R^2	.796	.740	.888	.839	.867	.611	.666
Adjusted R^2	.791	.728	.882	.832	.861	.605	.661

Notes: Constant included in the regressions but not shown on table. Standard deviation of coefficients between parentheses. Influential observations outside 2 standard deviations (Australia and New Zealand for 1870, Australia for 1890, Gabon and Venezuela for 1960).

To sum up, Temple and Johnson's results are confirmed. Moreover, regression results show that the observed associations between social development and income are significantly different from zero in all years discussed, and the coefficients are very similar between them. Thus, we can conclude that social development is positively related to economic wellbeing, and this relationship already existed in the nineteenth century.

4. II. 2. ENDOGENEITY AND CAUSALITY: MULTIVARIATE REGRESSION ANALYSIS

The main concern regarding the results in the previous section is whether the significant relationships we found can be attributed to a causal link between the two variables. In other words: can we conclude that social development significantly contributes to income being of a certain level, at least in part? This would be the case if we can assume that differences in the social development index between countries can be interpreted as exogenous changes. In our case, it is very unlikely that this is the case. On the contrary, it is very likely that there is some degree of endogeneity in the SDI, because the richer a country is the more likely to have an advantageous social quality thanks to being able to afford better education, higher likeliness to have advantageous political framework, etc. Therefore, before concluding that SDI has some partial causal effect on income one should control for all those factors that are known to affect income and could potentially be correlated with the social development index at the same time. The disturbance or error term in the regressions signify all the factors that play a role in

the determination of GDP that we have not taken into account. As a result, the regression estimate of the slope of SDI is likely to be biased, and can as a rule therefore not be interpreted as causal effects straight away (Wooldridge, 2001, chapter 1; Greene, 2003, chapter 8). If SDI is endogenous, then the estimate of its slope in the above equations cannot be interpreted as an estimate of the causal effect of SDI on GDP; it just verifies a statistical association between the two.

Graphs in the last section show a positively linear relationship between the two variables, but they do not represent a causal link *per se*. In order to find a causal effect between social development and income, we have to address the endogeneity of social development. The obvious solution to this problem is to include these omitted variables in the regressions. At this respect, there is a large literature on growth regressions from which we can borrow. The encompassing summary model by Bleaney and Nishiyama (2002) indicates that according to the most influential multivariate regressions literature, a comprehensive model should include openness to trade, life expectancy, schooling, quality of institutions, democracy, government savings, exports of primary products, climate and active population growth.

There are some data limitations to running multivariate regressions on nineteenth century information. For the last third of the nineteenth century, data are, in general, reliable (Mitchell, 1998). However, the specific control variables proposed in the growth regressions literature –representative examples are Barro (1997), Easterly and Levine (1997), and Sachs and Warner (1997)– and summarised by Bleaney and Nishiyama (2002) were not originally though for inclusion in pre-Second World War regressions. In many cases we are forced to find proxies available for the nineteenth century. Fortunately, Bleaney and Nishiyama (2002) summary of relevant variables is conceptually broad enough to facilitate this task. As we have an extensive list of variables in the database, we are able to find a proxy for all of these variables.

We use the following variables as proxy; (we present two alternative variables for some of them):

- foreign economic dependency ("foreignd") for openness to trade;[3]

[3] Growth of exports ("xgrgroup") from Morris and Adelman (1988) has also been alternatively tested for, giving similar results.

- proportion of population living in towns of more than 10,000 people ("urbani") for the inverse of life expectancy;[4]
- literacy rate ("lit") for schooling;
- predominant form of land tenure ("landtenu") for institutions;
- the index "polity2" from the Polity IV database or, alternatively, political influence of workers ("sociopol") for democracy; [5]
- relevance of the economic role of the government ("govt") for government savings;[6]
- shift in the export structure towards manufactured products ("shiftx") for exports of primary products; and
- Cumulative population growth in the last 20 years, percent, ("popgr") for active population growth.

All these variables are extracted from Morris and Adelman (1988), the Polity IV database (CIDCM) or derived from own calculations, all detailed in Chapter 2, and have the advantage that we have them for all countries in the sample.

- Finally, the proportion of land area in the geographical tropics ("tropicar") accounts for climate in the regressions. It has been shown that what really matters about climate for economic development is whether or not a country is located in the tropics (Sachs, 2001). We use the proportion of land area within the geographical tropics as in Gallup and Sachs (1999), and we take it as constant over the period of study. These geographical can be obtained from the *Geodata* database (Boston College, 2000).

So, <u>all</u> control variables suggested by Bleaney and Nishiyama (2002) are present in the analysis. This explains the very high R-squares obtained in the regressions. Table 4.2 displays the multivariate estimation results for 1870 and 1890.

[4] Life expectancy around 1900 could be taken from Maddison (1991 and 1995). This possibility has been explored; however it reduces the full sample size from 46 to 34, and any cross country cut from 23 to 17 countries only. Due to this limitation of the sample size, urbanisation has been preferred as a proxy. Results go along the same lines.

[5] Results are also shown for the alternative variable political power of workers ("sociopol") from Morris and Adelman (1988). It seems worth to bring this alternative in the table since it achieves larger significance results for 1890.

[6] Public expenditure as a percentage to GDP ("govrel") coming from own calculations detailed in Chapter 2 has also been tested. Results go in the same direction.

Table 4.2
Social Development and Income in the Late Nineteenth Century Including Control Variables.
Dependent variable: Log GDP Method:OLS

	(1)	(2)	(3)	(4)	(5)	(6)
Year	1870	1870	1890	1890	Pooled	Pooled
Observations	23	23	23	23	46	46
SDI	.421***	.390***	.450***	.429***	.427***	.401***
	(.063)	(.069)	(.119)	(.085)	(.048)	(.048)
Foreign dependency	-.0004	-.0005	-.002	-.004*	-.001	-.0015
	(.0015)	(.0014)	(.002)	(.002)	(.001)	(.0009)
Urbanisation	-.0007	-.0009	.002	.003	.0003	.0005
	(.0017)	(.0017)	(.002)	(.001)	(.0012)	(.0011)
Literacy rate	.0060***	.006***	.008***	.007***	.008***	.007***
	(.0016)	(.001)	(.002)	(.002)	(.001)	(.001)
Form of land tenure	-.006***	-.006***	-.006***	-.006***	-.007***	-.007***
	(.001)	(.001)	(.002)	(.001)	(.001)	(.0009)
Polity2	-.0008	-	.003	-	.004	-
	(.0059)		(.006)		(.004)	
Political power of workers	-	.002	-	.0044***	-	.003*
		(.002)		(.0019)		(.001)
Econ Role of Govt.	.002	.002	.002	.001	.002*	.002*
	(.002)	(.002)	(.002)	(.001)	(.001)	(.0009)
Shift of exports to manuf	-.0034*	-.0034*	-.005***	-.005***	-.005***	-.004***
	(.0019)	(.0018)	(.002)	(.001)	(.001)	(.001)
Population growth	.0037***	.0035***	.002	.001	.0030***	.003***
	(.0007)	(.0006)	(.001)	(.001)	(.0005)	(.0004)
% land in tropics	-.508***	-.521***	-.401***	-.341***	-.494***	-.465***
	(.109)	(.094)	(.125)	(.097)	(.072)	(.064)
R^2	.984	.985	.983	.988	.979	.981
Adjusted R^2	.971	.972	.968	.977	.973	.975

Notes: Constant not shown. Standard errors between parentheses. Test for heteroskedasticity of standard errors passed.*Coefficient statistically significant at the 0.1 level. ** Coefficient statistically significant at the 0.05 level. ***Coefficient statistically significant at the 0.01 level.

The dependent variable is again the log GDP per capita and the explanatory variables this time include the new Social Development Index (SDI) for the corresponding year plus the whole set of control variables discussed above. Table 4.2 presents 6 equations, 2 alternative ones for 1870, 2 for 1890, and 2 with a pooled sample. The regressions for individual years (regressions 1 to 4) are limited to a sample of 23 countries, those for which the Social Development Index is available for the nineteenth century. The pooled sample (regressions 5 and 6), has 46 observations. Still, adjusted R-squares are very high in all cases, around or above 0.98. This means that about 98 percent of the total international variation of income per capita is explained by the data.

The two equations for each year or pooled sample represent 2 alternative models depending on which variables we use as proxies. It is useful to show different proxy alternatives because in this way we see how much the regression results are affected by changing the variables of control. Table 4.2 reveals that all 6 equations have highly significant and very similar coefficients for SDI, all around 0.40 and 0.45. This means that 1 unit increase in the SDI, ranging approximately between -2 and 2, is associated with a 0.40-.45 unit increases in the log of income per capita. Data do not reveal any strong break between 1870 and 1890, which suggests that the relationship between social development and income enjoyed some generalised stability in the late-nineteenth century.

As a conclusion for this section we should say that the results show that, after controlling for a range of variables, the social development index SDI still has a significantly positive effect on income, and this relationship seems fairly stable for the period under analysis. The results for the nineteenth and the twentieth centuries are not directly comparable because we are not working with strictly the same set of countries and variables. However, putting together the findings for the twentieth century described in the introduction and ratified later on in the chapter, there are clear indications of a long run structural positive relationship between the SDI, understood as a proxy for social capital, and economic wellbeing.

4. II. 3. SDI AND FUTURE INCOME

As an extra piece of analysis, this section takes up on Putnam's (2000) hypothesis that social capital in a society does not have all its impact on wealth immediately but is something that transforms the society deeply and slowly enough to have effect several decades ahead of our time. This section presents some basic statistical results showing the increasing correlation between SDI and future income as we widen the time gap, which would ratify Putnam's hypothesis.

Thus, the next question is the following: What is the correlation between SDI and future income? Table 4.3 presents the simple univariate regression coefficients for SDI in 1960 and income per capita in different horizons: contemporary income per capita, income per capita 25 years ahead (emulating the Temple and Johnson 1998 chronology) and

income per capita 40 years ahead. The results show that the positive correlation between SDI and income becomes stronger the wider is the time gap. This means that the Social Development Index alone could predict the international distribution of income 40 years ahead better than the international distribution of income 25 years ahead, and much better than the international distribution of income for a contemporary time cut. This result is robust after controlling for the initial level of income. Once we account for the initial level of income this result does not change: A first regression of SDI on income is run, and then the residual SDI left unexplained by the initial level of income is taken to a second stage regression. Table 4.4 displays the results for the second stage, regressing future income on the residual SDI, which has been cleared of all the effects that the initial level of income in 1960 might have. Even after controlling for the effect of initial GDP, the SDI residual in 1960 alone is able to explain up to half of the variation in income per capita. Graphs 4.6 to 4.9 illustrate this relationship with all countries in the sample.

Let us now move to testing whether a positive relationship between SDI and future income existed in the past. The SDI residual for the nineteenth century has been constructed as a unique pooled variable for data corresponding to both time cuts, 1870 and 1890, in order to increase the sample size. However, these data do not correspond to a pooled regression making no difference between the two years. On the contrary, SDI residuals for the two time cuts correspond to different regressions run separately for the two years, which implies that the SDI 1870 residual has been extracted from contrasting it against 1870 per capita GDP only and, likewise, SDI 1890 residual arises from regressing SDI 1890 on per capita GDP for 1890 only. Therefore, every SDI residual corresponds to the regression against its relevant year. Just for reassurance, the whole exercise has been repeated with a pooled sample SDI residual, obtaining similar results.

Table 4.3 – SDI in 1960 and Future Income

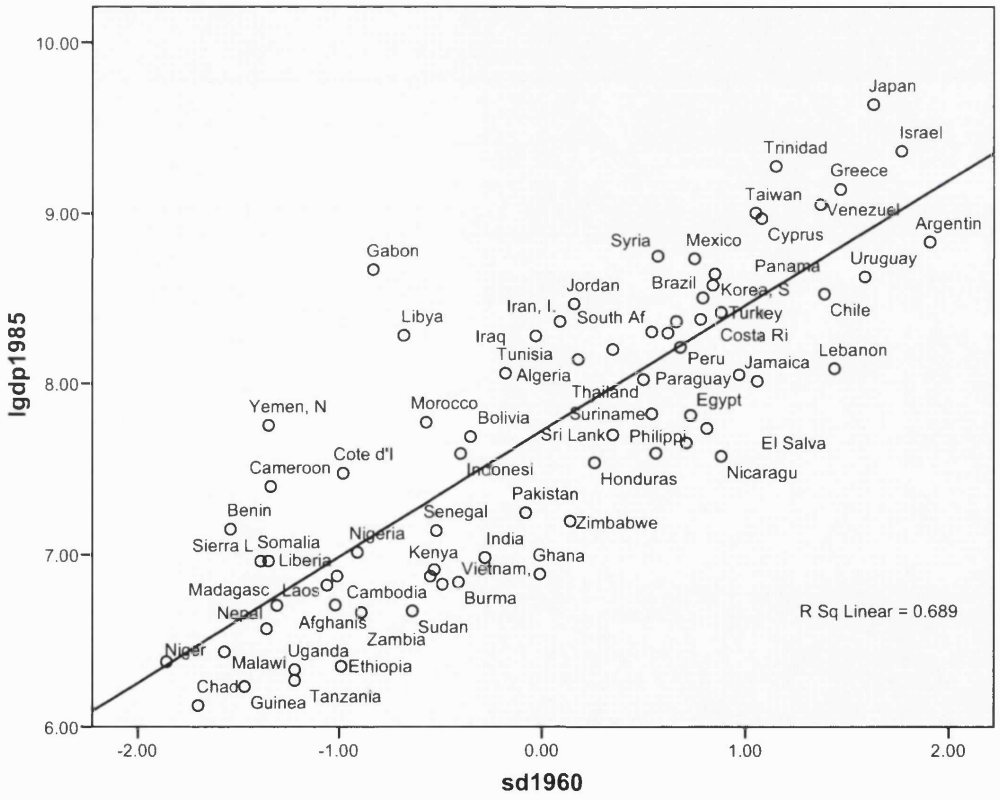
Regression	(1)	(2)	(3)
Dep. Var	Lgdp1960	Lgdp 1985	Lgdp2000
N obs.	74	74	72
Sd1960	.549***	0.733***	.855***
	(.052)	(.058)	(.060)
R-squared	.611	.689	.741
Adj R-sq	.605	.684	.737

Standard error in parenthesis. Constant not shown. Regression 3 includes 72 countries instead of 74 because comparable income per capita is not available for Cyprus and Suriname. ***Coefficient statistically significant at the 0.01 level.

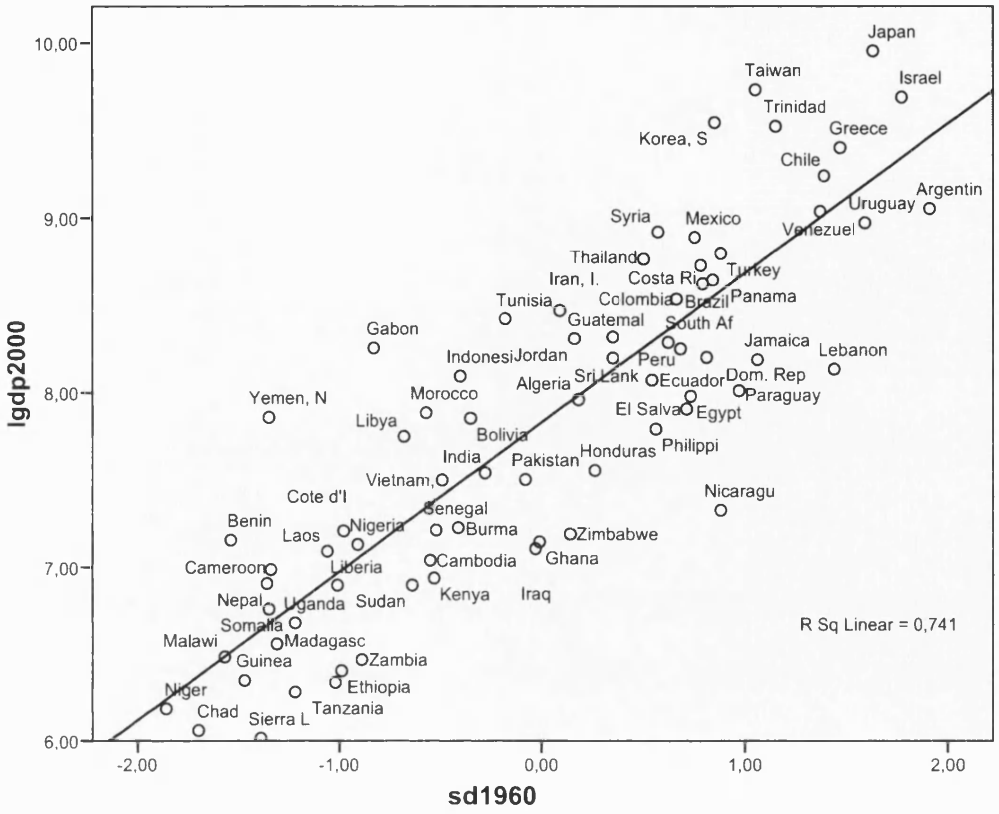
Table 4.4 – Residual SDI in 1960 and Future Income

Regression	(1)	(2)	(3)	(4)
Dep. Var.	Lgdp1985	Lgdp1985	Lgdp2000	Lgdp2000
N obs.	74	71	72	71
Sd 1960 residual, orthogonal to Initial GDP	.521*** (.127)	.824*** (.116)	.733*** (.134)	1.058*** (.120)
Standardised beta coefficient	.436	.650	.547	.733
R-squared	.190	.422	.299	.537
Adj R-sq	.179	.414	.289	.530

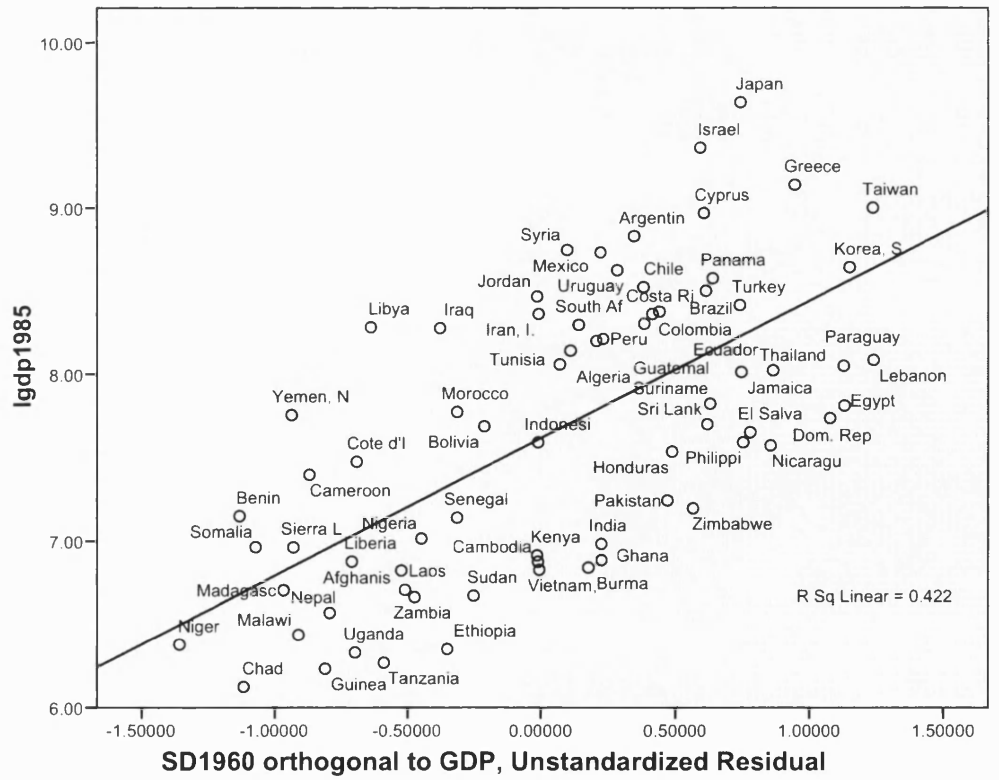
Standard error in parenthesis. Constant not shown. Regressions 2 and 4 exclude outliers above 2 standard deviations away from the mean (Gabon, Trinidad and Venezuela). ***Coefficient statistically significant at the 0.01 level.



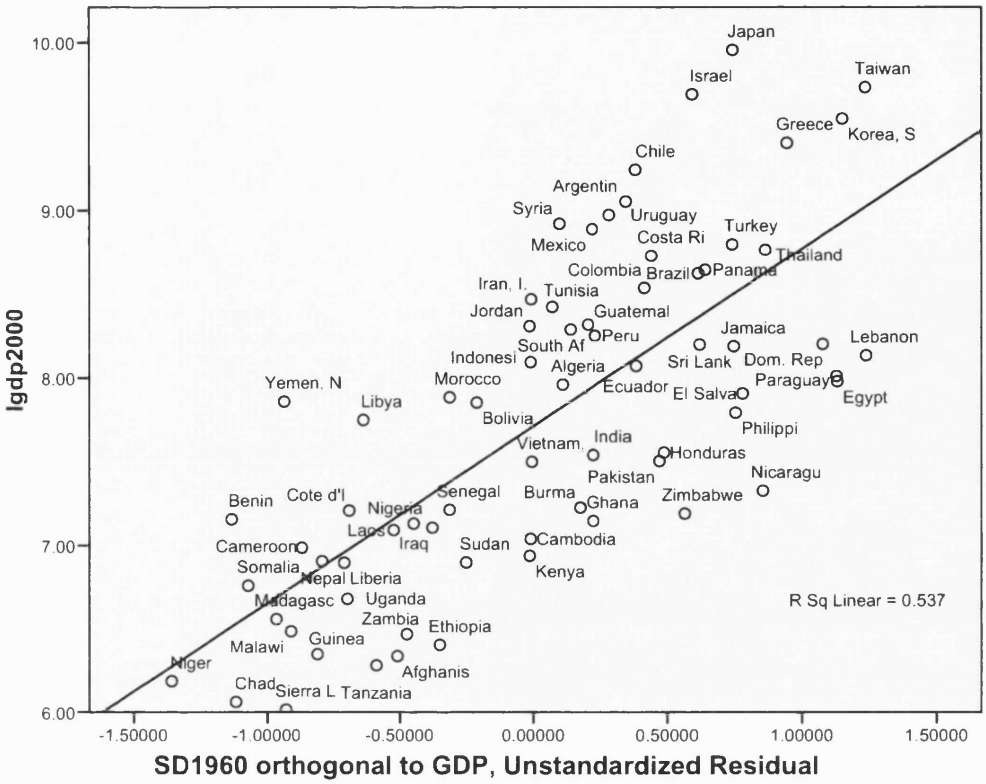
Graph 4.6 – Social Development and Income 25 years ahead



Graph 4.7 – Social Development and Income 40 years ahead



Graph 4.8 – Social Development Residual and Income 25 years ahead
Without Gabon, Trinindad and Venezuela (above 2 std dev away from the mean)



Graph 4.9 – Social Development Residual and Income 40 years ahead

Without Gabon, Trinindad and Venezuela (above 2 std dev away from the mean)

For the nineteenth century, no consistent relationship has been found between SDI and future income for 20 and 40 years ahead. This might be due to the fact that the economy simply moved slower 100 years ago, so one should expect to find a significant relationship between SDI and future income with wider time gaps than for the twentieth century analysis. This idea is backed up by the separate analysis of the two time cuts, 1870 and 1890. Despite the limitations in the sample size (23 countries only) the analysis for 1890 reveals some mild but positive correlation between SDI 1890 residual orthogonal to income and the distribution of per capita income 40 years ahead (equal to 0.108) above all other nineteenth century related correlations, which are close to 0. Further support of this idea is provided by Putnam (2000) historical study on the United States, where he argued that social capital at that time can have its stronger impact up to 70 years later.

A formal test for slower course of affairs (and, therefore, retarded impact) in the late nineteenth century with a gap of 60 or 70 years into the future is too daring since it involves including not only one but the two World Wars, which implied very abrupt disruptions both in the society and the economy. A preliminary survey for this

hypothesis at the international level with the information currently available does not give conclusive results.

4. III. THE RELATIONSHIP BETWEEN SDI AND ECONOMIC GROWTH

The previous sections investigated the relationship between the SDI and income *per se*. This section explores the relationship between the SDI and future long run economic growth.

The long run growth rate is different from the annual growth rate. The annual growth rate is the change in GDP over one year. The long run growth rate is the overall change in GDP over a period longer than one year, and it is generally computed as the difference in logs on income between 2 more or less distant years.[7] Given the long run character of the potential effects of social capital, it would make no sense to study the short run one year effect of the overall level of the SDI in a country at a given point in time. What makes more sense is to study the potential effect of social capital into the economic growth of a country over a prolonged period of time. Many studies on long run growth rates, including those cited in the introduction, are based on 20-25 years periods. For instance, Temple and Johnson (1998) use the log-difference of per capita GDP between 1985 and 1960 as the dependent variable in Figure II above.

The objective is to detect how much of the growth to come could have been predicted by just looking at the SDI at the beginning of the period. Because SDI at the beginning of the period is correlated with GDP at the beginning of the period, we need to control for this fact prior to the analysis. In general, economic theory predicts a negative correlation coefficient between the rate of growth and initial GDP. Countries that start from low levels of GDP tend to grow faster, and vice versa. This is called the "catching-up" effect (Abramovitz, 1986). In order to counteract the catching-up effect, we follow the same procedure as Temple and Johnson (1998): In a first stage, a linear regression of SDI on Log *per capita* GDP is run. Now, the resulting residuals of this regression are

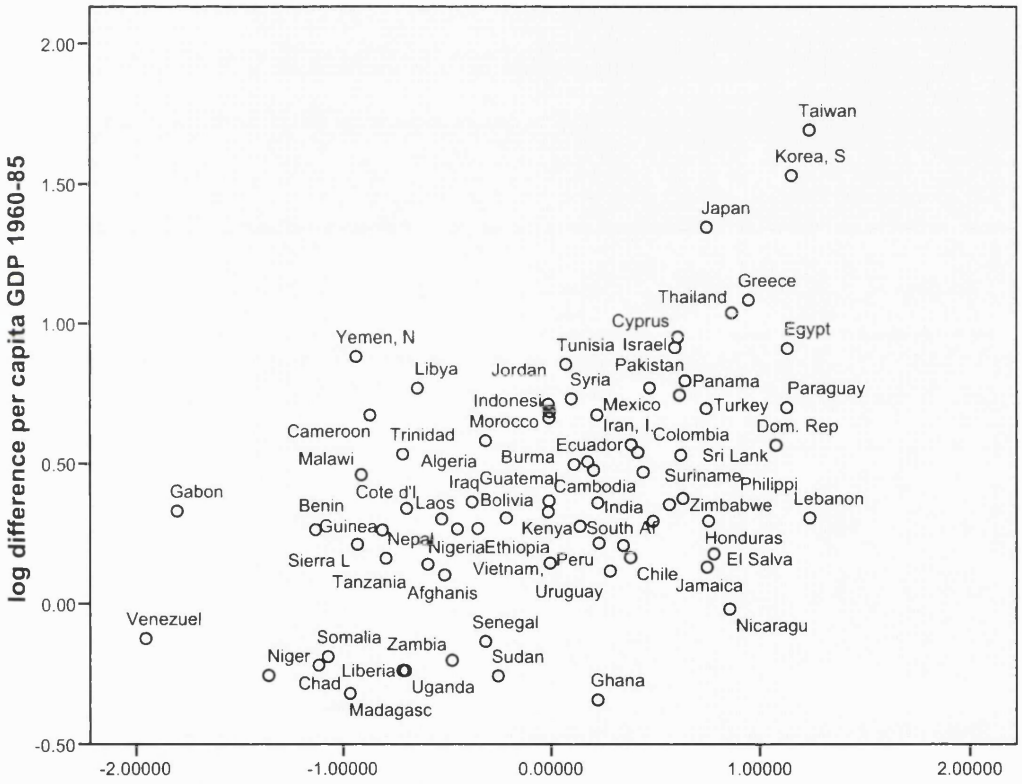
[7] Generally, the long run growth rate between two designated years is approximated by the log-difference on per capita GDP. For instance, the long run per capita growth rate for the period 1870-1890 is: GR187090=(PCGDP1890–PCGDP1870)/PCGDP1870. Notice that GR187090=(PCDGP1890/PCGDP1870)-1 is a small positive number. Given that the logarithm is a good approximation of a function for any small positive number, this expression can be approximated: GR187090=LPCGDP1890–LPCGDP1870. This is the log-difference (difference in logarithms). Growth regressions literature has generally used this formulation.

the part of SDI that is not affected by income. On a second stage, we run the regression we are really interested in, which is the long run rate of growth on the orthogonalised SDI residual. This procedure was used by Temple and Johnson (1998:972).

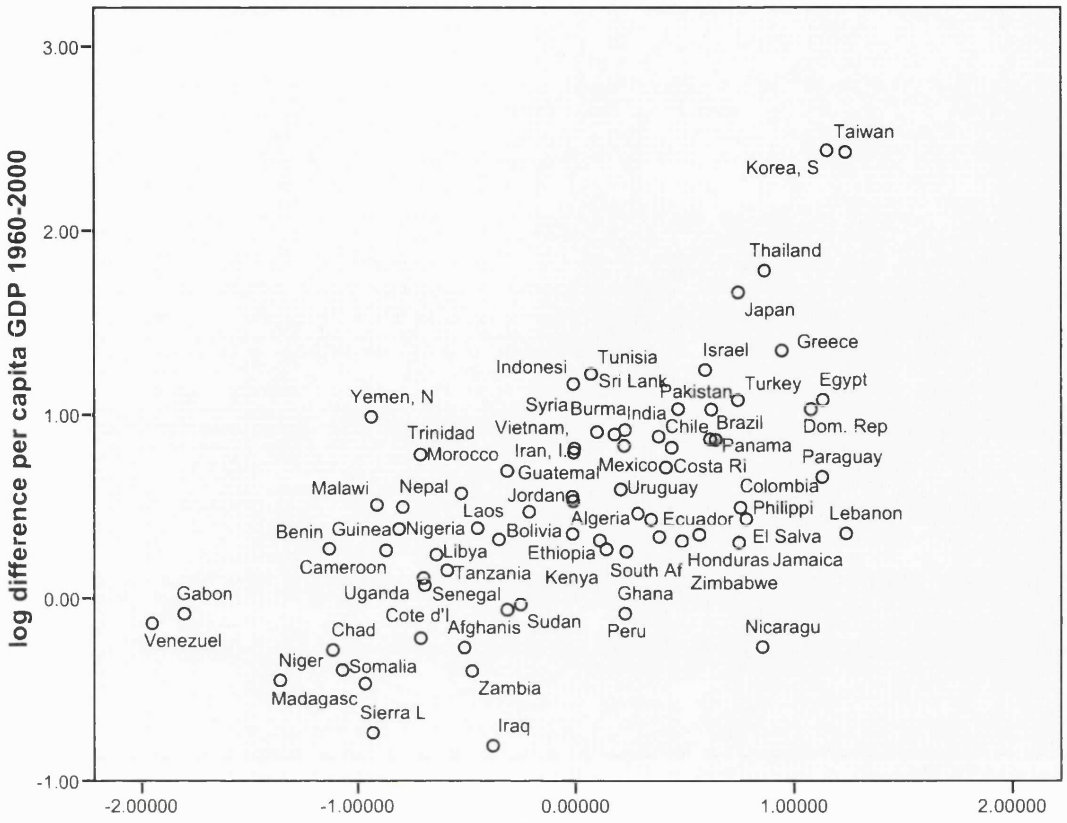
Long run growth rates have been calculated for the following three periods: First, the more recent 1960-1985 –as in Temple and Johnson (1998)– and, then, for the earlier chronologies, basically corresponding to the periodisation established by Morris and Adelman (1988), 1870-1890, 1890-1910. Then, these are plotted against the Social Development Index (SDI) for the initial year of each period (1960, 1870 and 1890, respectively) which is orthogonal to initial income. As in the previous section, this is done through running a first stage regression of SDI on income, and then using the residual in a second stage set of regressions as the part of the index that cannot be explained by factors simultaneously correlated with income per capita. Finally, some additional longer periods have been taken into account as well, mainly 40-years period, 60-years period, etc. Future long run growth for this longer time spans have been calculated and contrasted with the residual of initial SDI orthogonal to *per capita* GDP at the beginning of the period. Results for the three SDI cross-sectional series available are presented in Tables 4.5, 4.6 and 4.7, corresponding to long run potential impact of the general assessment of the quality of a society in 1960, 1870 and 1890, respectively.

Graph 4.10 depicts the correspondence between the long run economic growth for the period 1960-1985 and the SDI residual orthogonal to initial income at the beginning of the period. It reveals a positive association between the two. On the other hand, Graph 4.11 corresponds to a longer time period of 40 years, accounting for accumulated economic growth for the period 1960-2000. The same positive association is found for the more prolonged time period, if not stronger. The latter would imply that the social mapping of the world around 1960 has a stronger correspondence with the world distribution of economic growth in the longer period 1960-2000 than with the shorter period 1960-1985. A formal statistical test for this relationship is presented in Table 4.5. The second stage regression of long run growth 1960-85 against the SDI component orthogonal to income reveals a coefficient of the magnitude of 0.298, which rises to 0.520 when the regression is run for the longer time period 1960-2000 (both statistically significant at the 99 percent confidence level). This means that 1 unit increase in the SDI corresponds to 0.298 more units of log-difference in income for a 25 year period, and 0.520 more units of log-difference in income for a 40 year period. The standardised

beta coefficients, which express the correlation between the dependent and the independent variables in terms of standard deviations, deliver magnitudes of 50 to 60 percent increase of a standard deviation in accumulated economic growth for every one-standard deviation increase in the social index. The beta coefficient for the period 1960-85 is 0.527 (above 50 percent of a standard deviation increase in long run growth for every 1 standard deviation increase in SDI), while the beta coefficient for the period 1960-2000 is 0.619 (above 60 percent of a standard deviation increase in long run growth for every 1 standard deviation increase in SDI at the beginning of the period).



Graph 4.10 – Social Development in 1960 and Economic Growth 1960-1985



Graph 4.11 – Social Development in 1960 and Economic Growth 1960-2000

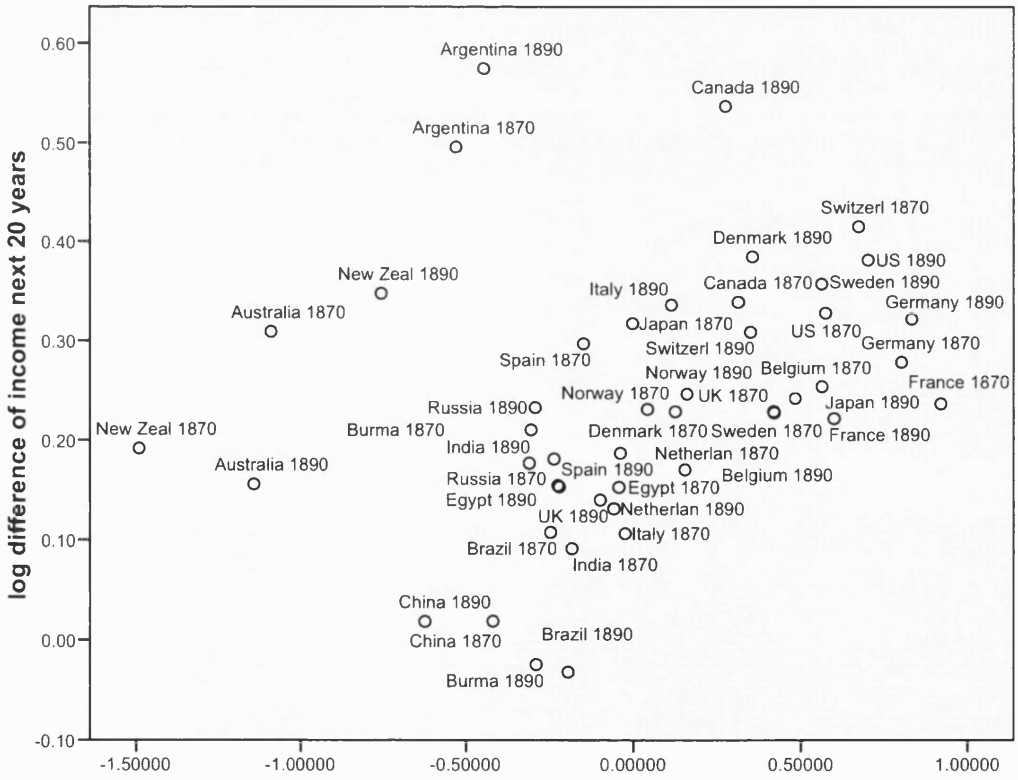
Table 4.5 - SDI 1960 and Future Long Run Growth
Dependent variable: Long Run Log- difference of Per Capita GDP

Period	1960-1985	1960-2000
N obs.	74	72
SDI 1960 residual, orthogonal to initial GDP	.298*** (.057)	.520*** (.079)
Std beta coeff.	.527	.619
R-squared	.278	.383
Adj R-sq	.268	.374

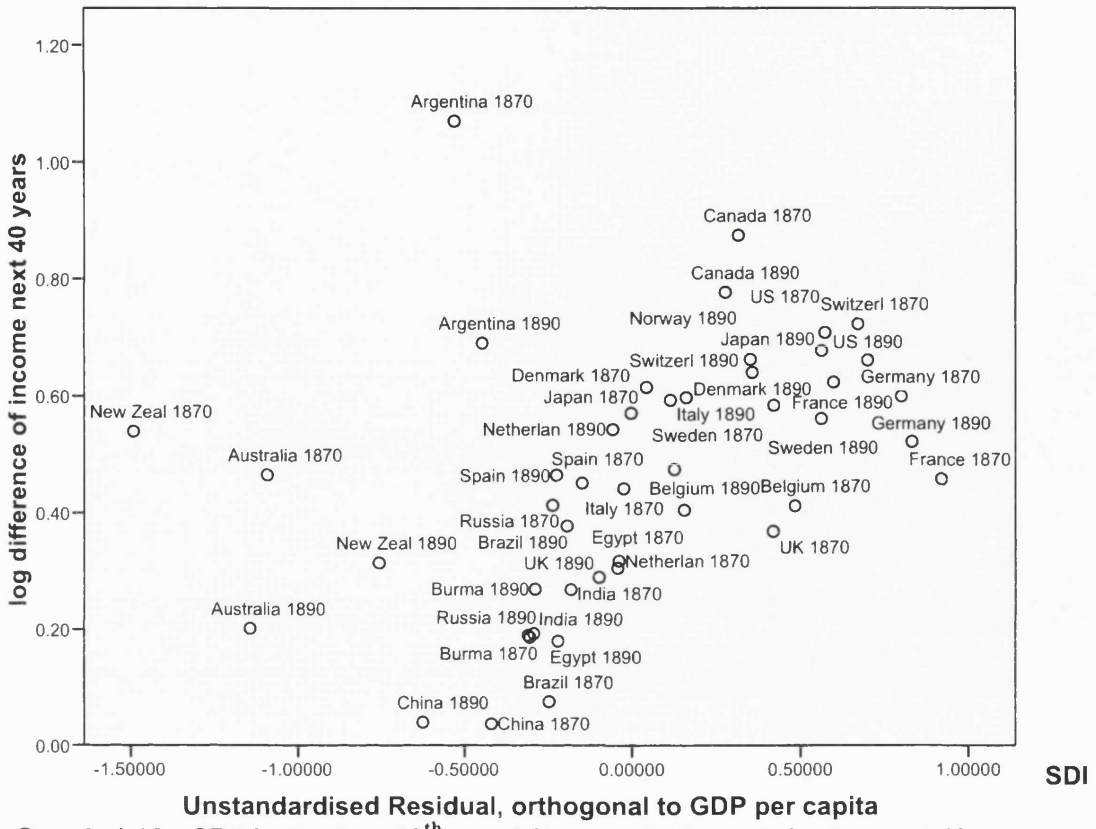
Notes: Constant not shown. Standard deviation of coefficients between parentheses. Influential observations outside 2 standard deviations. *** Coefficient statistically significant at the 0.01 level.

Can one find similar results for the nineteenth century? The same test with the new nineteenth century SDI allows us to test whether a persistent relationship between social capital and subsequent long run growth already existed at that time. Thus, the next question is the following: Would a social capital proxy for the nineteenth century have

been able to say something about subsequent economic growth? The answer is yes. There is some evidence for a stable relationship between nineteenth century SDI and subsequent long run growth. Graph 4.12 plots a pooled sample of late nineteenth century SDI for all countries available against the accumulated long run growth in the next 20 years, corresponding with the 20 years period breaks established by Morris and Adelman (1988). As before, the SDI has first been cleared up of any effect that *per capita* income could have. Given the restricted sample of countries available for the nineteenth century both measures for 1870 and 1890 have been plotted together in the graphs. This does not seem unreasonable given the fact that social changes are slow and the two years, 1870 and 1890, are not too far apart in time. Furthermore, plotting them together gives a comparative perspective of performance. Notice, however, that SDI residual has been calculated independently for both years, with separate regression weights; this is, that every point has been calculated using their



Graph 4.12 - SDI in the late 19th c and Economic Growth in the next 20 years



Graph 4.13 - SDI in the late 19th c and Economic Growth in the next 40 years

corresponding year, using two separate regressions for the two years. The observations are shown together in a unique graph as a matter of presentation, not of calculation. In any case, formal statistical tests are calculated separately for the two years and shown in the tables below.

Table 4.6 - SDI1870 and Future Long Run Growth
Dependent variable: Long Run Log- difference of Per Capita GDP

Period	1870-1890	1870-1910	1870-1928	1870-1960
N obs.	23	23	23	23
Sd1870 residual, orthogonal to initial GDP	.035 (.040)	.074 (.090)	.214** (.101)	.321* (.175)
Standardised beta coeff.	.188	.177	.420	.371
R-squared	.035	.031	.176	.138
Adj R-sq	-.011	-.015	.137	.097

Notes: Constant not shown. Standard deviation of coefficients between parentheses.
** Coefficient statistically significant at the 0.05 level.
*** Coefficient statistically significant at the 0.01 level.

Table 4.7 - SDI1890 and Future Long Run Growth
Dependent variable: Long Run Log- difference of Per Capita GDP

Period N obs.	1890-1910 23	1890-1910 22	1890-1928 23	1890-1928 22	1890-1960 23	1890-1960 22
Sd1890 residual, orthogonal to initial GDP	.109 (.064)	.143** (.054)	.292*** (.066)	.325*** (.058)	.523*** (.156)	.473*** (.135)
Standardised beta coeff	.350	.509	.695	.782	.591	.618
R-squared	.122	.259	.483	.611	.350	.382
Adj R-sq	.080	.222	.459	.592	.319	.351

Notes: Constant not shown. Standard deviation of coefficients between parentheses. Influential observations outside 2 standard deviations excluded in 2nd regression (Argentina for 1890-1910, 1890-1928; Burma for 1890-1960). ** Coefficient statistically significant at the 0.05 level.
*** Coefficient statistically significant at the 0.01 level.

What are the findings? The late nineteenth century SDI residual orthogonal to income, like its twentieth century counterpart, does also show a positive correlation with future economic growth. This stands clear from both Graphs 4.12 and 4.13, depicting this relationship for the next 20 and 40 years respectively.

Tables 4.6 and 4.7 summarise the results of a formal statistical analysis separately for the two years 1870 and 1890. The results of this analysis should be interpreted within the limitations that the small sample size of 23 countries imposes. There is no clear statistical evidence for correlation between the SDI in 1870 and future long run economic growth for the coming 20 or 40 years. Only when one looks at longer time spans does significant positive correlation arise, of the order that was discussed above for the twentieth century case. The coefficient of the SDI residual after orthogonalising to initial *per capita* GDP, is of the magnitude of 0.214 for a 50 years growth period and 0.321 for a 90 years growth period, both statistically significant. The standardised beta coefficients run around 40 percent, comparable to 50 to 60 percent results for the twentieth century. More precisely, the standardised beta coefficient for the future 50 years growth regression is 0.420, which implies that rising SDI by one standard deviation is associated with 42 percent increase of a standard deviation in accumulated long run growth.[8] Finally, for a very long run period, running from the first SDI available corresponding to 1870 to the most recent for 1960 (adding up to a total of 90 years accumulated growth), the beta coefficient is 0.371, quite close to the 50 years

[8] Notice that 1928 GDP figures where used in the regressions instead of 1930 in order to avoid the Great Depression.

period. This invites to think that almost all the gains from social development status had already been realised in the five-decade period, or at least started declining somewhere in between.

In order to explore this idea of increasing and then declining correlation after a certain number of decades further, let us have a look at the results for 1890 SDI, shown in Table 4.7. The coefficients for the long run growth regressions start becoming significant earlier than for the 1870 case (for shorter, closer time periods) and then remain significant throughout. This fact taken together with the even shortest-period significance in Table 4.5 for the twentieth century reveals that correlations with economic growth for earlier periods take longer to appear. This makes sense if we think that economic progress has accelerated greatly during the course of the twentieth century, and things simply happen faster. The historical empirical evidence presented here supports this hypothesis. SDI takes 50 years to report significant coefficients for long run economic growth in 1870, 40 years for 1890 and 20 years for 1960. Of course, these numbers are subject to the arbitrariness of the time periods chosen and would need a lengthier and more costly exploration to underpin when is exactly the turning point at which the social index for a certain year turns significant for future long run economic growth. Given the limitation of the sample size, it makes no sense to pursue a study of this kind with the information available. That would be an attempt to push data too far.

Detailed regression results for 1890 are to be found in Table 4.7, which shows the second stage regressions, with the long run log-difference of per capita GDP as the dependent variable for different periods and the residual of a first stage regression of initial SDI against per capita GDP as the independent variable. The coefficient of the residual of 1890 SDI after orthogonalising to GDP runs from 0.109 when we look at its linear association with accumulated economic growth in the next 20 years, rises to 0.292 when we increase the growth period to 40 years and finally reaches 0.523 when we look at the next 70 years (accumulated growth up to 1960), in a clear steadily increasing tendency. Because of the small sample size of 23 countries, results are especially susceptible of being affected by influential observations. In order to test for this, regressions for all three considered periods have been re-run excluding influential observations defined as being more than 2 standard deviations away from the sample mean. This excludes one individual from each of the regressions: Argentina for the nineteenth century regressions and Burma for the twentieth century ones. Re-running

the regressions without the influential observations yields the following coefficients: 0.143 for the regression on future 20 years accumulated economic growth, 0.325 for the regression on future 40 years and 0.473 for the regression on future 70 years, up to 1960. All coefficients are statistically significant and the steadily increasing tendency of the slope is confirmed. So, the further into the future we project the level of social development the steeper the linear relationship becomes. As far as explanatory power is concerned, adjusted r-squares are maximum for the medium long-run period of 40 years, reaching almost 60 percent with the residual SDI as only explanatory variable. The standardised beta coefficients reach its maximum in this time span too, reaching 78.2 percent co-movement in terms of standard deviations for the period 1890-1928 without Argentina. The average for all periods goes more on the lines of 50 to 60 percent association, coinciding once more with the coefficients of the twentieth century regressions.

Overall, in all three cases, for the 1870-onwards analysis, for the 1890-onwards and for the 1960-onwards one, the slope of the fitted line is steeper for longer time periods. At the same time, it is also the true that the standardised beta coefficients first increase with the length of the time period accounting for accumulated economic growth, pointing at long run effects of social development; and then decrease after 5-6 decades when enough historical perspective is observable. Furthermore, the regression coefficients are comparable for the three SDI cross-sections under study, the earlier chronologies taking longer to manifest in the form of statistically significant positive association. This fact could have some economic meaning, implying that things simply move faster for more recent chronologies.

4. IV. CONCLUSION

This is an empirical investigation on the economic welfare implications of social capital for an international panel of countries, with special emphasis on the late-nineteenth century and achieving a comparative perspective with the twentieth century. This chapter presents some evidence pointing at a stable long run structural relationship between social capital and macroeconomic performance. Overall, we can conclude that social development is positively related to economic wellbeing, and this relationship already existed in the nineteenth century.

Firstly, it is confirmed that in the post-Second World War era countries with higher levels of social development are associated with higher levels of income and vice versa. Secondly, making use of a new social development index for 1870 and 1890, it is found that a positive relationship between social development and log per capita income already existed in the late nineteenth century. At this respect, a strong positive linear association between the two variables is found. Regression results show coefficients that are significantly different from zero in all years discussed, and very similar between them. This upholds after controlling for foreign trade volume and structure, urbanisation, education, quality of institutions, political stability, government expenditure, population growth, and climate.

On the other hand, a statistically significant relation between the social capital proxy and future long run economic growth for the three social development cross-sections available starting in the late-nineteenth century is also found. The coefficient (or slope) associated with initial SDI increases when we account for longer periods of accumulated economic growth, suggesting that the association between social development and economic growth takes several decades to unfold completely. Social changes do not manifest its strongest impact in the economy immediately; they need several decades to unfold their transformations into economic outcomes. Indeed, both twentieth and nineteenth century statistical analysis reveal a stronger correspondence of the world distribution of SDI with the world distribution of income and growth for longer rather than shorter time horizons. The level of association seems to be reaching its maximum level after 5-6 decades and then starts declining again. Nevertheless, the maximum association horizon seems to be shrinking the closer we get to the present time. This suggests that events simply seem to be happening faster the closer we are to the present. The magnitude of the association in the maximum correlation periods ranges around 50 to 60 percent of a standard deviation increase in accumulated long run growth for every 1 standard deviation increase in SDI. This means that looking at the SDI at the beginning of the period would have been able to say quite a lot about the future world distribution of economic growth.

REFERENCES

Abramovitz, Moses, 1986 June, 'Catching Up, Forging Ahead, and Falling Behind', *Journal of Economic History*, 46 (2):385-406.

Adelman, Irma, and Cinthia Taft Morris, 1967, *Society, Politics, and Economic Development: A Quantitative Approach*, John Hopkins Press (Baltimore).

Barro, Robert J., 1997, 'Determinants of Economic Growth: A Cross-Country Empirical Study', MIT Press (Cambridge, Massachusetts).

Bleaney, Michael and Akira Nishiyama, 2002, 'Explaning Growth: A Contest Between Models', *Journal of Economic Growth*, 7: 43-56

Boston College, Faculty Microcomputer Resource Center (FMRC), 2000, Graduate Statistical Assistant Program, *Geography and Economic Development.* Access to *Geodata.*

Center for International Development and Conflict Management (CIDCM), *Polity IV Project* http://www.cidcm.umd.edu/polity/ (last accessed 27/11/2007).

Easterly, William, and Ross Levine, 1997, 'Africa's Growth Tragedy: Policies and Ethnic Divisions', *Quarterly Journal of Economics*, 112 (4): 1203-50.

Gallup, John Luke, and Jeffrey D. Sachs, with Andrew Mellinger, 1999 March, 'Geography and Economic Development', CID working paper 1, http://www2.cid.harvard.edu/cidwp/001.pdf

Goldberger, Arthur S., 1991, *A Course in Econometrics*, Harvard University Press.

Greene, William H., 2003, *Econometric Analysis*, Pearson Education (Upper Saddle River, NJ).

Heston, Alan, Robert Summers and Bettina Aten, 2006, September, *Penn World Table Version 6.2*, Center for International Comparisons of Production, Income and Prices at the University of Pennsylvania.

Knack, Stephen, and Philip Keefer, 1997, 'Does Social Capital Have an Economic Payoff? A Cross-Country Investigation', *The Quarterly Journal of Economics*, 112(4):1251-1288.

La Porta, Rafael, Florencio Lopez-de-Silanes, Andrei Shleifer, and Robert W. Vishny, 1997, 'Trust in Large Organisations', *American Economic Review Papers and Proceedings*, 87 (2):333-338.

Maddison, Angus, 1991, *Dynamic Forces in Capitalist Development: A Long-Run Comparative View*, Oxford University Press (Oxford and New York).

—— , 1995, Monitoring the World Economy 1820-1992, OECD Development Centre (Paris).

—— , 2003, *The World Economy: Historical Statistics*, OECD Development Centre (Paris).

—— , 2007, August, *World Population, GDP and Per Capita GDP, 1-2003 AD*, Angus Maddison Homepage http://www.ggdc.net/maddison/ (last accessed 11/04/2008).

Mankiw, N. Gregory, David Rome, and David Weil, 1992, 'A Contribution to the Empirics of Economic Growth', *The Quarterly Journal of Economics* CVII:407-437.

Mitchell, Brian Redman, 1998, *International Historical Statistics: Europe 1750-1993*, Fourth edition, Stockton Press and MacMillan Reference Ltd, (New York and Basingstoke respectively).

Morris, Cynthia Taft and Irma Adelman, 1988, *Comparative Patterns of Economic Development, 180-1914*, The John Hopkins University Press (Baltimore and London).

Putnam, Robert D., 1993, *Making Democracy Work: Civic Traditions in Modern Italy*, Princeton University Press, (Princeton, New Jersey).

Sachs, Jeffrey D., 2001, 'Tropical Underdevelopment', NBER working paper 8119, http://www.nber.org/papers/w8119

Sachs, Jeffrey D., and A. Warner, 1997, 'Sources of Slow Growth in African Economies', *Journal of African Economies*, 6: 335-376.

Summers, Robert, and Alan Heston, 1988, 'A New Set of International Comparisons of Real Product and Price Levels Estimates for 130 countries, 1950-85', *Review of Income and Wealth* XXXIV:1-26.

Temple, Jonathan, and Paul A. Johnson, 'Social Capability and Economic Growth', *The Quarterly Journal of Economics* 113 (3), 1998, pp. 965-990.

Wooldridge, Jeffrey M., 2001, *Econometric Analysis of Cross Section and Panel Data*, MIT Press.

Chapter 5

AN EMPIRICAL APPLICATION OF THE SOCIAL DEVELOPMENT INDEX INTO ECONOMIC THEORY

INTRODUCTORY NOTE TO CHAPTER 5:

After all attention devoted to social capital research and findings in recent years, can one go on without incorporating social capital into economic theory? Not any more; social capital should not be neglected in economic theory research any more. Chapter 5 of the thesis provides an example of a practical application of the social capital indicators into a contemporary economic geography investigation.

How SDI Can Help Testing the Market Access Matters Hypothesis in Historical Perspective

ABSTRACT

The Redding and Venables (2004) series for Market Access for 1994 are replicated in order to re-estimate the GDP equation with alternative socio-institutional measures. Using the newly created series together with more standard socio-institutional indicators allows testing where does cross-country variation in the performance of the Market Access index arise from. The results show that the Redding and Venables (2004) results are not unconditional. Poor governance undermines the positive benefits of geographical location. Only when a country enjoys high enough institutional quality can it effectively benefit from accessibility to markets. In a second part, Market Access is calculated for other benchmark years. Using the newly created series together with the proposed social capital proxies allows testing of the relationship between market access and GDP for another benchmark year. The results show that the relationship between access to markets and GDP was as strong in the mid-1960s as it has been for the mid-1990s, advocating no death of distance.

5. I. INTRODUCTION

Under increasing world economic integration, why firms do not move more production to low wage countries? Some factors previously studies are natural endowments or technology related arguments. Redding and Venables (2004) try to answer this question by incorporating geographical location into the analysis.

The mechanism in which they focus is distance to markets. Both distance to inputs (capital and intermediate goods) and distance to output (final production) markets are considered. Under given technology and internationally set prices (except wages), firms located further away from markets bear extra costs to trade that force wages downwards in order to remain competitive. This mechanism would explain why there are not more firms moving to low-wage countries.

Geographic location determines wages. Redding and Venables (2004) find statistically significant effects of geographical location on per capita income, after controlling for primary resource endowments, and a number of institutional characteristics. The magnitude of these effects is important. Halving a country's distance to markets would result in about a 25 percent increase in per capita income (Redding and Venables, 2004:77). But, can one generalise this result? Moving countries that are very far away from the main world markets into a central location can result into much larger income effects. For instance, performing the experiment of hypothetically moving Zimbabwe to central Europe results into an almost 80 percent rise in Zimbabwe's GDP per capita (Redding and Venables, 2004:77). But, can one really move any country in the world to Europe and make it work?

This chapter explores the insights of the Redding and Venables (2004) results, and shows that they are non-robust and particularly sensitive to the institutional setup. An institutional quality threshold is established below which no matter where we are in the world the hypothetical rise in GDP per capita is not going to happen.

5. II. MODEL

The theoretical framework used by Redding and Venables (2004) follows the general equilibrium model exposed in Fujita, Krugman, and Venables (1999), chapter 14 on international specialisation. This model allows for international specialisation with intermediate goods. The final manufactured good is also used as an input in the production function, thus, acting as well as an intermediate capital good. Production also requires an immobile (non-tradable) factor of production, which can be interpreted as labour.

In the exposition of the model, we follow Fujita, Krugman, and Venables (1999), and introduce some notation changes to match Redding and Venables (2004) application. For instance, we allow for i = 1,..., R countries instead of two in the original Fujita, Krugman, and Venables (1999).

Prices of production factors are w_i for labour and G_i for the intermediate good. When the latter is sold directly to the consumer its price is p_i. These define the indirect production function

$$p_i = w_i^{1-\alpha} G_i^{\alpha}, \qquad 0< \alpha <1 \qquad (1),$$

which is Cobb-Douglas with intermediate manufactured good share α. Equation 1 illustrates the fact that firms set price equal to marginal cost.

In each country i, there are n firms producing n differentiated manufactured products. Therefore, n_i is the number of varieties of the manufactured good produced in country i. The manufactured good enjoys a constant elasticity of substitution (CES) amongst all its varieties. The CES function is

$$G_i = [\sum_{j=1}^{R} n_j (p_j T_{ji})^{1-\sigma}]^{\frac{1}{1-\sigma}}, \qquad \sigma >1 \quad (2),$$

where T_{ji} stands for the transportation costs from country j to country i. Firms chose to buy all varieties available to produce at internal price G_i, and the more varieties the

better. Equally, consumers get best utility by purchasing all varieties available, and the more the better. Their CES utility function is

$$U_j = [\sum_{i=1}^{R} n_i x_{ij}^{\frac{\sigma-1}{\sigma}}]^{\frac{\sigma}{\sigma-1}} \qquad (3),$$

where x_{ij} is the amount of the manufactured good produced in country i which is demanded by county j. In other words, x_{ij} represents the level of exports from i to j, and the level of internal demand in the case of j = i. Aggregating across importing countries we have

$$\sum_{j=1}^{R} x_{ij} = x_i \qquad (4),$$

where x_i is the aggregate level of production for a given firm-variety in country i.

How much is the expenditure of country i on manufactured goods? If we define Y_i as income in country i and μ as the share of manufactures that go to final consumption, then total expenditure of country i on manufactured goods, E_i is equal to the sum of consumers' demand plus intermediate good's demand on behalf of producers.

$$E_i = \mu Y_i + \alpha n_i p_i x_i \qquad (5)$$

In equation 5, μY_i is the proportion of income that goes to direct consumption of manufactures and $\alpha n_i p_i x_i$ is the proportion of total production that is devoted to the purchase of intermediate goods. Notice that α is the Cobb-Douglass share of the tradable input and x_i represents the equilibrium level of production. Therefore, $n_i p_i x_i$ is the value of production of country i, which we can denote by X_i. So,

$$E_i = \mu Y_i + \alpha X_i \qquad (6)$$

X_{ij} is the value of exports from country i to country j. We will later focus on this variable (value of exports).

Now, we want to know the number of varieties n_i. In order to simplify the model, we follow Fujita, Krugman, and Venables (1999) by setting an arbitrary breakeven point of sales ($\bar{x}$)[1]. The breakeven point of sales is the same for every country because they enjoy the same technology. Then, this breakeven point determines the salaries.

$$\bar{x} = \frac{1}{1-\alpha} \qquad (7),$$

Choosing the breakeven point equal to $\frac{1}{1-\alpha}$ simplifies the calculation. Equation 7 implies that

$$(1-\alpha) n_i p_i \left(\frac{1}{1-\alpha} \right) = w_i \lambda_i \qquad (8)$$

so

$$n_i p_i = w_i \lambda_i \qquad (9)$$

and, therefore,

$$n_i = \frac{w_i}{p_i} \lambda_i \qquad (10)$$

So, the number of varieties in every country is proportional to the real wage. The higher the real wage, the larger the number of industrial varieties. Likewise, the larger the share of the labour force in manufactures, the larger the number of industrial varieties.

In order to obtain price equations for the intermediate good, we incorporate n_j and p_j into G_i equation. First, we incorporate the resulting equation for n_j, equation 10, into G_i, equation 2;

$$G_i = [\sum_{j=1}^{R} \frac{w_j}{p_j} \lambda_j (p_j T_{ji})^{1-\sigma}]^{\frac{1}{1-\sigma}} \qquad (11)$$

[1] Fujita, Krugman, and Venables (1999) use a different notation for $\bar{x}$. They use q* for the level of sales at the zero-profit equilibrium instead (see Fujita, Krugman, and Venables, 1999:242). Here we choose $\bar{x}$ notation in order to make it match with that of Redding and Venables (2004).

and, then, we incorporate the indirect production function, equation 1, determining the price of the consumption good, p_j, as a function of the prices of inputs, w_j and G_j.

$$G_i = [\sum_{j=1}^{R} \frac{w_j}{w_j^{1-\alpha} G_j^{\alpha}} \lambda_j (w_j^{1-\alpha} G_j^{\alpha} T_{ji})^{1-\sigma}]^{\frac{1}{1-\sigma}} \qquad (12)$$

Rearranging it renders

$$G_i^{1-\sigma} = \sum_{j=1}^{R} \lambda_j w_j^{1-\sigma(1-\alpha)} G_j^{-\alpha\sigma} T_{ji}^{1-\sigma} \qquad (13)$$

which is the price equation for country i.

The structure of the price equation is the same as in Fujita, Krugman, and Venables (1999) and Redding and Venables (2004), but, unlike them, we initially allow for a higher number of countries, R. The referred previous models consisted of a world of two countries only.

On the producers' side, firms follow a profit maximising behaviour. Profits are as follows:

$$\pi_i = p_i x_i - w_i (F + c x_i) \qquad (14)$$

where F represents the fixed costs of production and c the variable costs, being the latter proportional to the quantity produced, q_i. Production is given by the demand function. Firms take the price of the intermediate input imported from another country, G_j, as given. x_i is equal to the proportion of total sales that go to final consumption. This amount is given by equation 15.

$$x_i = \mu \sum_{i=1}^{R} Y_i (p_i T_{ij})^{-\sigma} G_i^{\sigma-1} T_{ij} \qquad (15)$$

Under a perfectly competitive environment, firms set price equal to marginal cost. This zero-profit condition gives rise to the optimal production choice for the firm, $\bar{x}$. Therefore, rearranging terms, the demand function is

$$\bar{x} = \mu \sum_{i=1}^{R} Y_i p_i^{-\sigma} T_{ij}^{1-\sigma} G_i^{\sigma-1} \qquad (16)$$

Now we can isolate the price of the consumption good, p_i.

$$p_i^{\sigma} = \frac{\mu}{\bar{x}} \sum_{i=1}^{R} Y_i T_{ij}^{1-\sigma} G_i^{\sigma-1} \qquad (17)$$

Elasticity of demand is σ.

$$p_i(1-\frac{1}{\sigma}) = cw_i \qquad (18)$$

or

$$p_i = cw_i \left(\frac{\sigma}{\sigma - 1} \right) \qquad (19)$$

Equation 19 is the pricing rule. The demand function and the pricing rule give rise to the wage equations. The next step is finding the wage equations. Applying the pricing rule to the inverse demand function found in equation 17 renders

$$cw_i \left(\frac{\sigma}{\sigma - 1} \right) = \left(\frac{\mu}{\bar{x}} \sum_{j=1}^{R} Y_j T_{ij}^{1-\sigma} G_j^{\sigma-1} \right)^{\frac{1}{\sigma}} \qquad (20)$$

Thus, the initial wage equation is as follows:

$$w_i = \frac{\sigma - 1}{c\sigma} \left(\frac{\mu}{\bar{x}} \sum_{j=1}^{R} Y_j T_{ij}^{1-\sigma} G_j^{\sigma-1} \right)^{\frac{1}{\sigma}} \qquad (21)$$

Equation 21 gives the wage at which firms in country i break even.

In order to simplify the calculations, we can do the following normalisations, without loss of generality:

$$F \equiv \frac{\mu}{\sigma} \qquad (22)$$

and

$$c = \frac{\sigma - 1}{\sigma} \equiv \rho \qquad (23)$$

By setting the fixed and variable costs of manufacturing production equal to certain parameters of our interest, we will be able to simplify the pricing rule and the wage equations. Thanks to the normalisation in equation 22, the pricing rule in equation 19 becomes

$$p_i = w_i \qquad (24)$$

and the wage equations in 21 become

$$w_i = \left(\sum_{j=1}^{R} Y_j T_{ij}^{1-\sigma} G_j^{\sigma-1} \right)^{\frac{1}{\sigma}} \qquad (25)$$

In equilibrium, the supply capacity of country i, s_i is

$$s_i = n_i p_i^{1-\sigma} \qquad (26)$$

If we weight the internal price by the iceberg transportation costs from country i to country j, the resulting expression will be expressed in terms of the price of domestically produced goods placed at the foreign market j. Adding up over all countries we get the resulting supplier access of country j.

$$SA_j = \sum_{i=1}^{R} n_i (p_i T_{ij})^{\sigma-1} \qquad (27)$$

Moving to the exports' market, the market capacity of country j, m_j, is defined as

$$m_j = E_j G_j^{\sigma-1} \qquad (28)$$

and the corresponding market access of country i, MA_i, is defined as the sum of all market accesses across countries, expressed in terms of the price once the good is placed in country j. i.e. weighted by the iceberg transportation cost from country i to country j.

$$MA_i = \sum_{j=1}^{R} E_j (G_j T_{ij})^{\sigma-1} \qquad (29)$$

X_{ij} is defined as the value of exports from county i to country j.

$$X_{ij} = n_i p_i x_{ij} \qquad (30)$$

Then, the value of exports gives rise to the following trade equation:

$$n_i p_i x_{ij} = n_i p_i^{1-\sigma} T_{ij}^{1-\sigma} E_j G_j^{\sigma-1} \qquad (31)$$

which, in terms of the above definitions, can be written as

$$X_{ij} = s_i T_{ij}^{1-\sigma} m_j \qquad (32)$$

5. III. ECONOMETRIC SPECIFICATIONS

This section goes from the theoretical model to econometric specifications; it deals with the issue of how to approach econometric specification for empirical estimation of the Market Access and Supplier Access indicators.

First, we take logarithms at both sides of the trade equation.

$$\ln X_{ij} = \ln s_i + (1-\sigma) \ln T_{ij} + \ln m_j \qquad (33)$$

Following Redding and Venables (2004), the supply capacity of the exporting country is estimated with exporting country characteristics (cty_i). The importing partner market capacity is estimated with importing country characteristics (ptn_j). The transportations costs between the two countries are estimated with the distance between capitals ($dist_{ij}$) and a common border dummy ($bord_{ij}$).

$$\ln X_{ij} = \theta + \mu_i cty_i + \lambda_j ptn_j + \delta_1 \ln dist_{ij} + \delta_2 bord_{ij} + u_{ij} \qquad (34)$$

In equation 34, all explanatory variables are dummy variables but distance between capitals; u_{ij} is the error term. Considering that, by nature, trade data are censored at 0, we prefer a Tobit estimation over ordinary least squares.

The Market Access and Supplier Access indicators are defined as follows:

$$MA_i \equiv \sum_{j=1}^{R} m_j T_{ij}^{1-\sigma} \qquad (35)$$

$$SA_j \equiv \sum_{i=1}^{R} T_{ij}^{1-\sigma} s_i \qquad (36)$$

and, according to the econometric specification in equation 34, they are calculated as

$$\ln \hat{MA}_i = \sum_{j=1}^{R} (\hat{\lambda}_j ptn_j + \hat{\delta}_1 \ln dist_{ij} + \hat{\delta}_2 bord_{ij}) \qquad (37)$$

$$\ln \hat{SA}_j = \sum_{i=1}^{R} (\hat{\mu}_i cty_i + \hat{\delta}_1 \ln dist_{ij} + \hat{\delta}_2 bord_{ij}) \qquad (38)$$

Now, the following step is to calculate the market and supplier access indicators empirically.

5. IV. EMPIRICAL DATA ESTIMATION

Redding and Venables (2004) estimated the Market Access and Supplier Access equations using 1994 data and then contrasted them to 1996 GDP per capita.[2] The NBER-UN International Trade Data on-line archive provides bilateral trade data for a given year since 1962 and until 2000. All bilateral trade data available for 1994 were extracted. This on-line archive is an updated version of the Canadian Statistics trade data archive. There is a change of approach in the new version. The latter compile information primarily from the importer side instead of from the exporter side, since this is supposed to be more reliable (Feenstra et al. 2005).

The Redding and Venables (2004) market access indicators for 1994 were replicated using their method[3]. Full series of foreign and domestic market access and supplier access were generated for a cross section of countries around the world. Total market access is calculated by adding up foreign and domestic market access. Given that the interest of this paper is to account for the effects of access to markets, the focus of the rest of this paper is on market access rather than supplier access.

Redding and Venables (2004), actually, did not use the NBER-UN trade database in their calculations, but an adapted version of these data, cleaned up by the CEP. The next step was to get hold of the CEP database they actually used. The CEP World Trade Database incorporates some changes described in detail in Stewart (2001). These lead to the existence of two alternative datasets: on the one hand, the original UN international trade dataset as the NBER presents it and, on the other hand, the CEP revised version. Although calculations are done for the two alternative databases to check robustness, the CEP cleaned up version of the bilateral trade data is preferred, because of the refinements it incorporates and because it is the one used in the Redding and Venables' study. The appendix contains the full new series created and the differences between the two alternative sources listed country by country. The world distribution of foreign market access generated by the two alternative datasets can be found in the final maps. Because the final impact in the results is not considerable, it is preferable to operate

[2] 1994 Market Access indicators are contrasted with 1996 GDP per capita in order to avoid simultaneity bias.

[3] Thanks to Stephen Redding, who passed me the original STATA code used in Redding and Venables (2004).

with just one database, the CEP one, for the reasons outlined above. However, some results using the NBER database can be found in the appendix.

The rest of the paper will make use of the market access series derived from the CEP database. GDP series are taken from the Penn World Table version 6.1 for consistency with Redding and Venables (2004), and geography related control variables are taken from their same source for the same reason.

5. V. HOW SDI CAN HELP TESTING MARKET ACCESS ROBUSTNESS IN A HISTORICAL FRAMEWORK

Bilateral trade data are available since 1962. It seems an interesting question to ask whether the relationship between Market Access as defined by Redding and Venables (2004) and per capita income context was more robust under an earlier historical context. Going back to the first NBER-UN International Trade Data reporting year gives 3 decades of perspective, and thus could turn into an interesting exploration.

Chronologically speaking, the beginning of the bilateral trade database coincides with the Social Development Index created for the early 1960s. Thus, the Social Development Index can become a candidate for acting as a proxy for social capital at that point in History. The SDI attempt to capture how well a society works together as a whole, which could be a proxy for social capital in the absence of data sourcing from surveys, such as the World Value Surveys, which started being conducted only in the 1970s. It is true that data coming from surveys, such as trust, and a compound index of historical statistics, such as the SDI, are not directly comparable, but they may, however, make reference to the same concept of social capital. Methodologically speaking, even the current indicators of good governance released by the World Bank (2005) were originally calculated as an extraction of a principal components analysis of socio-institutional data, which is the same procedure with which the SDI was calculated. So, the use of the SDI as a proxy for social capital and placing it in the context of the socio-institutional control variable in the regressions above does seem reasonable. If one is willing to take the SDI as our institutional indicator for the 1960s, the test performed in previous sections can be repeated here for the time corresponding to the beginning of the bilateral trade database. Even more so if we take into account the

explanatory power the SDI has revealed in previous chapters, which would by itself justify the potentiality of the results in the following econometric tests.

We performed a similar test of market access influence on GDP making use of the social development index available for the 1960s (Adelman and Morris, 1967) as means of controlling for social capital, as suggested by Temple and Johnson (1998). Luckily enough, the initial reporting time of the bilateral trade data coincides with that of the Adelman and Morris' Social Development Index, which opens room for testing around this time. New series of market access indicators for the beginning of the 1960s were created using the Redding and Venables (2004) method. The new Market Access series together with the use of alternative socio-institutional indicators (the Social Development Index), allows testing of the economic geography theory for another benchmark year. We used the newly created 1965 market access series, together with the Adelman and Morris (1967) social development index and the falciparam malaria index, which is available for 1966 from the Center for International Development (CID) at Harvard University. Purely geographical control variables do not differ from those used for 1996.

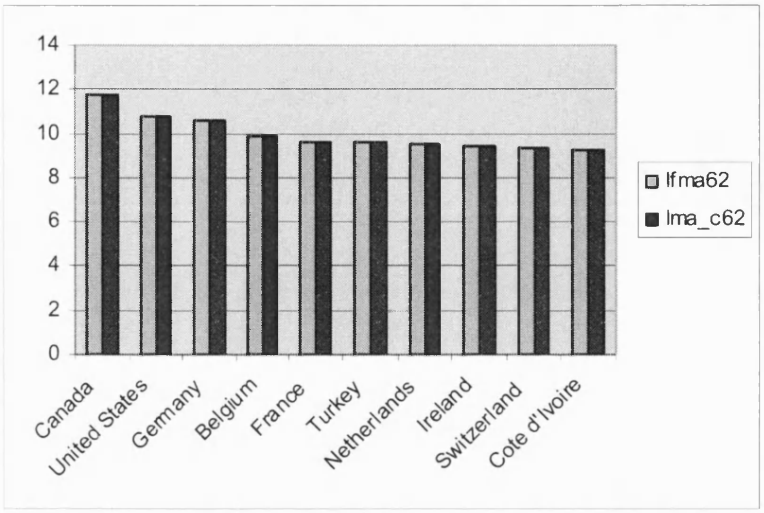
5. VI. HISTORICAL DATA ESTIMATION

In previous sections, we replicated Redding and Venables (2004) market access indicators for 1994 using their method. Now we created new series of market access indicators for other benchmark years using the same method. Market access series can be calculated for every year between 1962 and 2000, since bilateral trade data are available from the NBER-UN International Trade Data for every year within this range.

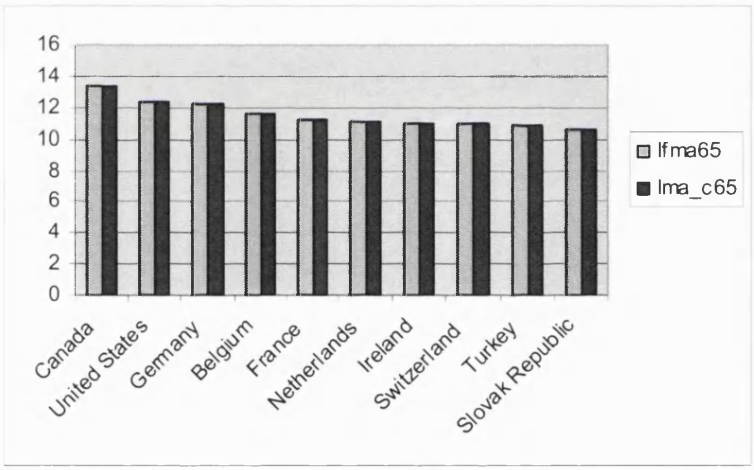
The following graphs illustrate the oscillations in levels and rankings of market access over the years; (based on own calculations, replicating Redding and Venables 2004 method).

lfma = log of foreign market access

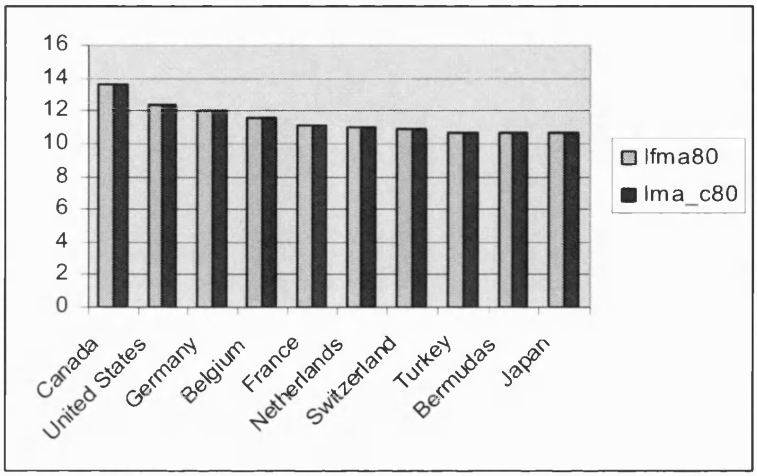
lma_c = log of market access MA(3), using Tobit estimation, as in Redding and Venables (2004)



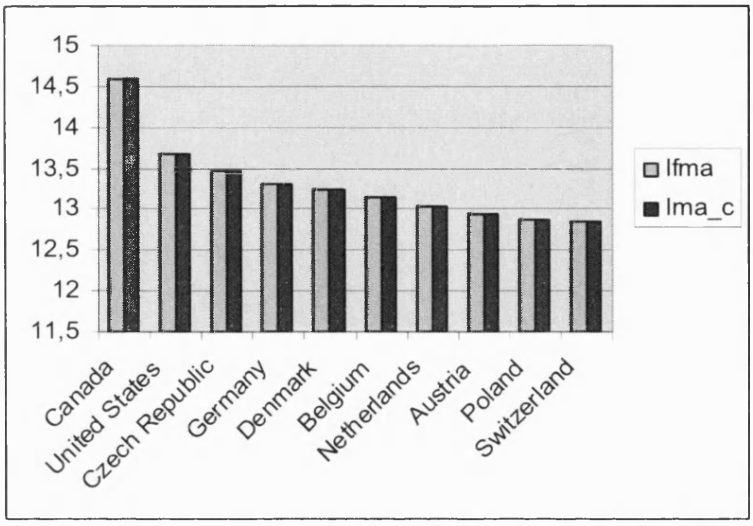
Ten Highest Market Access Scores for 1962 (bilateral trade data initial year)



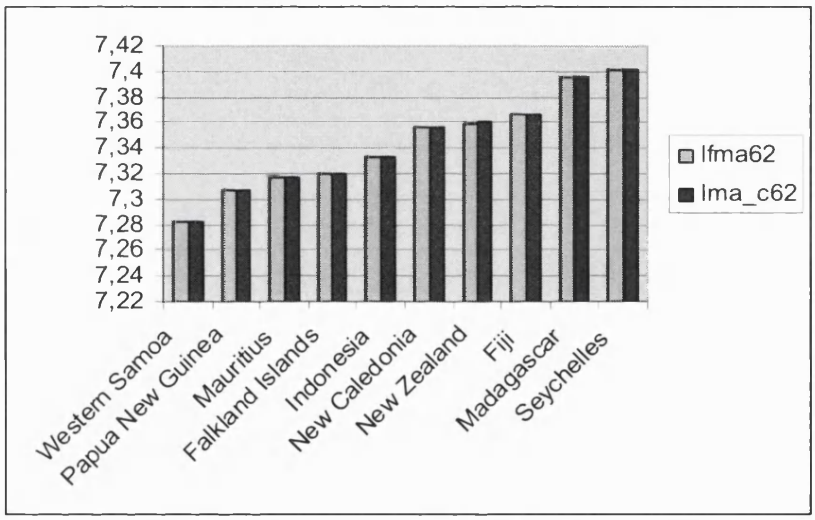
Ten highest Market Access Scores for 1965 (year used for testing)



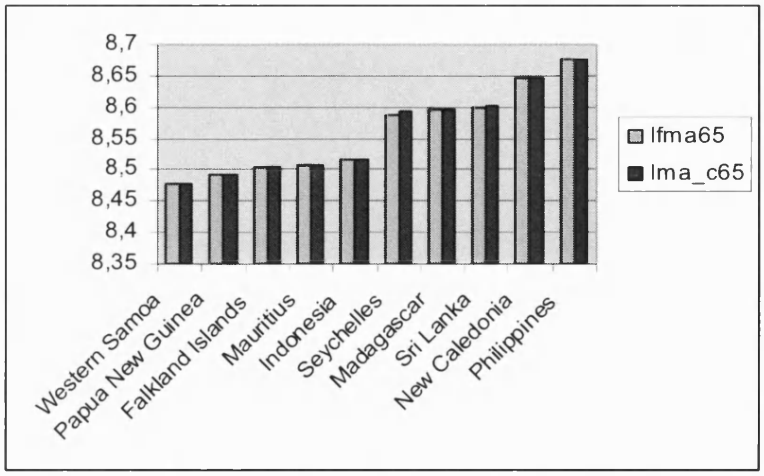
Ten Highest Market Access Scores for 1980 (intermediate year)



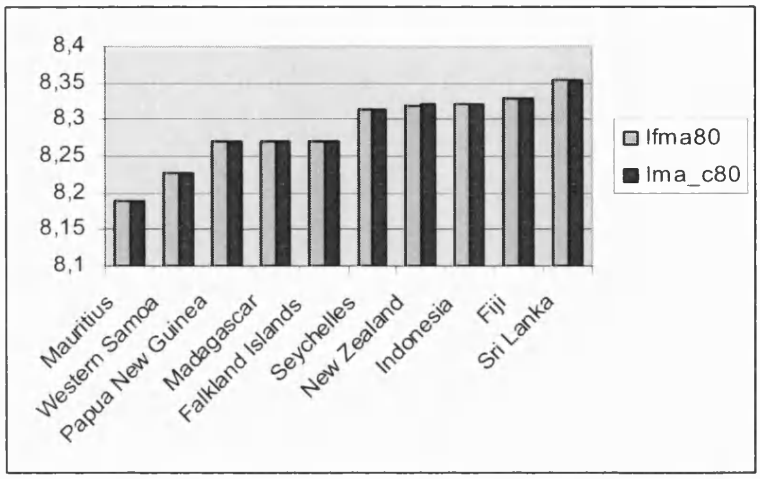
Ten Highest Market Access Scores for 1994 (year used for testing by Redding and Venables 2004)



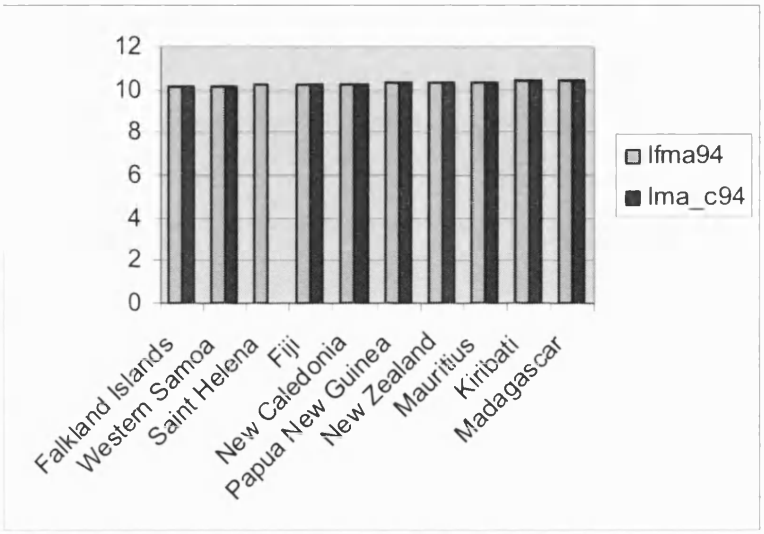
Ten Lowest Market Access Scores for 1962 (bilateral trade data initial year)



Ten Lowest Market Access Scores for 1965 (year used for testing)



Ten Lowest Market Access Scores for 1980 (intermediate year)



Ten Lowest Market Access Scores for 1994 (year used for testing by Redding and Venables 2004)

5. VII. TESTING THE THEORY FOR ANOTHER BENCHMARK YEAR

After obtaining the Market Access indicators for several years, we can now re-test the theory of whether access to markets matters for GDP for alternative benchmark years under the context of the 1960s, for which the SDI is available[4]. Table 5.1[5] displays the results.

[4] Recall that the Social Development Index constructed by Adelman and Morris (1967) uses data ranking from 1959 to 1962. Therefore, the Index is a good indicator for the beginning of the decade of the 1960s starting in 1962 and also valid for contrasting with data corresponding with successive years. Testing should never be done retrospectively, since this could reverse the direction of causality.

[5] Tables in this chapter are displayed at the end because of format restrictions.

Similar regressions to those run in Redding and Venables (2004:69) have been run focusing on the log of real GDP per capita in 1964 and 1965 (to test several years shortly after 1962). Regressions with 1964 GDP do not pass the joint significance F-test; this might be due to simultaneity bias. Therefore, it is preferable to run the regressions on 1965 GDP, which poses no statistical problem. Approximately 67 percent of the disparities in 1965 real GDP per capita are captured in our economic geography regressions (only 9 percent below Redding and Venables' explanatory power for the 1990s with only 2/3 of the observations). Also Redding and Venables (2004)'s R-squares for the 1990s diminish with sample size. For instance, our R-squares for 1965 are 9 percent smaller of what Redding and Venables could explain with a sample size of 91 for 1994 trade data; but slightly higher than what they could explain with a sample size of 69. Regarding the focus economic geography variable (Market Access), the regression coefficients do not reach statistical significance, which might be due to the limitations of the sample size. Nevertheless, the sign of the coefficient is consistently positive across all specifications, which suggests that better access to markets is positively associated with higher levels of GDP per capita.

Overall, given the data availability, the relevance of the market access indicators cannot be proven for 1965 as strongly as for 1994 in this section. The next section explores this topic further. The next question is: Which aspects of the market access indicators do matter? In the next section, we explore the impact of plain distance to markets thanks to an instrumental variables' test.

5. VIII. DOES MARKET ACCESS MATTER THROUGH DISTANCE? INSTRUMENTAL VARIABLES' TEST

This section presents some evidence of the fact that bilateral distances may have played some role in the development of the World trade flows. What is the relationship between the Market Access indicators developed by Redding and Venables (2004) and distances to the main World markets? Do Market Access indicators matter just because they capture distance to markets? An instrumental variables test would help us answer this question in the two chronological setups, the 1990s and the 1960s. And more interestingly, it is possible to perform an equivalent distance-instrumental variables' test for both periods if the use similar control variables.

We ought to question whether distance is important directly by itself (the further away a country is the more unlikely it is to supply it with domestic goods and services in a proportional way), or indirectly through the more complex market access indicators (which include other aspects such as costs). If market access matters mainly through distance, we would be able to instrument the market access indicators with distance and then re-run the main regressions getting significant coefficients. Redding and Venables (2004) successfully instrumented their 1994 market access variables with distance to the centre of the main World markets, (which they considered Belgium, Japan, and the United States). In this section, we show that distance to the centre of the three main World markets can successfully instrument the complex market access indicators. We perform the instrumental variables (IV) test with 1994 trade data, and then check the robustness of the results by performing another IV test for our other benchmark year, 1965.

The instrumental variables test takes the form of a 2-stages least squares (2SLS) estimation procedure, where the first stage instruments the Market Access indicators with distance to the three main World market centres; namely USA, Belgium and Japan; and an extra variable which is the minimum amongst the three. The forth instrument is important because it captures whether being close to just one of the maim markets is enough to boost bilateral trade. Then, the second-stage estimation uses the Market Access indicators as instrumented in the first stage instead of the originals. In this way, it is possible to discern whether access to markets matters through distance, which does indeed seem to be the case according to the results presented herein. Finally, this procedure is repeated for the two historical benchmarks, the 1990s and the 1960s.

Table 5.2 shows the result of the instrumental variables estimation for the mid-1990s. Market Access data have been calculated for 1994, replicating Redding and Venables (2004) procedure, contrasted to 1996 log of real GDP per capita from the Penn World Tables as done in previous sections. Moreover, Market Access data have also been calculated directly for 1996, since we should not expect simultaneity bias problems with GDP this time because the access to markets is instrumented with purely exogenous variables (distances). Both alternatives are presented in Table 5.2, showing very similar results. The Market Access indicators consistently achieve significantly positive coefficients for all alternatives. Its magnitude is around 0.30 when using 1994 trade data (see specifications (1) and (2) in Table 5.2) and 0.29 when using 1996 trade data (see

specifications (3) and (4) in Table 5.2), above the 0.215 achieved by Redding and Venables (2004:69) reproduced here in Table 5.4, specification (1). Thus, it does seem that when instrumenting the Market Access indicators with distance, they are able to explain real GDP per capita differentials across the World. In one word, location matters. Being far away from the potential trading partners does make things difficult.

We have been using bilateral distance between capital cities throughout ("distcap" in the CEPII database). Alternative distance definitions: "dist" (distance between the most populated cities), "distw" (arithmetic weighted average distance, weights by population in several cities within each country), and "distwces" (weights according to CES function) give similar results. For a detailed description of the four distance measurement alternatives, see Clair et al. (2004:4). For a discussion of advantages and inconveniences, see Head and Mayer (2002). Though differences in results are small, "distwces", the measure preferred by Head and Mayer, brings sometimes slightly higher R-squares and statistical significance of the coefficients in our regressions.

The next question is whether the location matters hypothesis can also be tested for another benchmark year in order to gain some historical perspective. Making use of the bilateral trade database we can resort to the initial reporting year 1962, as done in the previous section, and apply the instrumental variable tests to the 1960s data. The result is shown in Table 5.3. Table 5.3 presents astonishingly similar results to Table 5.2. Again, Market Access coefficients recover statistical significance when instrumented by distance to main markets. Their sign is positive and their magnitudes are comparable to Redding and Venables (2004:69), again staying above. Using 1962 trade data gives rise to 0.33 significant coefficients at the 5 percent significance level; using 1965 trade data produces 0.31 significant coefficients at the same significance level. So, the level of association of access to markets instrumented via bilateral distances and real GDP per capita was as strong in the 1960s as it has been shown to be in the 1990s, and there is no sign of 'the death of distance' from these regressions (phrase making reference to the book *The Tyranny of Distance*, by Geoffrey Blainey on Australian history (Blainey,1966)).

5. IX. ROBUSTNESS TESTS WITH MORE STANDARD SOCIO-INSTITUTIONAL VARIABLES

In this section, the robustness of the effect of access to markets on GDP is explored further. This is done by means of replicating the results with alternative socio-institutional control variables.

Regression (1) in Table 5.4 corresponds to Redding and Venables (2004); regression (2) is my replica[6]. The results go in the same direction, although there are some observed discrepancies of unknown nature. These are the baseline equations. Regressions (3), (6) and (8) in Table 5.4 have been run using the two-stages least squares procedure described in the previous section as an alternative to the ordinary least squares estimation. If we run the regressions using the logarithm of the distance to the main world markets as done in the previous section, we will resort to the original Redding and Venables' results: The coefficients of Foreign Market Access turn out to be positive and statistically significant, their magnitudes are of the order of 0.20 to 0.30, recovering the Redding and Venables's results. So it seems that the market access indicators recover significance if instrumented through the exogenous variable distance. Thus, market access matters indirectly via distance to the main world markets, pointing at the hypothesis that geography still matters.

The rest of the equations in the table illustrate how changes in the choice of socio-institutional variables alter the sign of the key variable (access to markets) and do not lead to a consistently significant relationship with per capita income. The sign of foreign market access changes depending on whether we run the regression on risk of expropriation (regression 2), or property rights (regression 4). These two variables are conceptually opposed and therefore, should be capturing the same effect, although with reversed sign. The fact is that including one or the other reverses the sign of the key variable (foreign market access). The sample is much smaller under "risk of expropriation", so the phenomenon could be due to sample selection bias. "Property rights" seems to be the one Redding and Venables (2004) actually used, because of the sample size exact coincidence. It has a wider sample than "risk of expropriation"; however, the latter covers the same concept and enjoys a much more detailed scale, so it

[6] I show 4 significative ciphers throughout, independently of the number of decimals. This is why, in some cases, equation (1) shows only 3 ciphers while the rest of equations tend to show 4 ciphers.

could be considered preferable if there were no sample size restrictions. Because of the wider sample size "property rights" is preferred.

In regressions (5), (6), (7) and (9) the risk of expropriation and property rights variables have been substituted by the Index of Economic Freedom, which is a wider concept that includes the former. This index gives strongly significant coefficients, but the sign of foreign market access still depends on which other socio-institutional control variables do we include. The truth is that the variables "socialist rule" and the war dummy (Redding and Venables's choice) return non-significant coefficients in all instances. So, what happens if we run the same regression with more standard socio-institutional variables?

Regressions (7), (8) and (9) incorporate the concept of social capital into the regressions. This is done through the variable "trust" as it appears in Knack and Keefer (1997) and has been extracted from the World Value Surveys[7]. The use of this variable has been suggested by Knack and Keefer (1997) and used widely ever since. It measures the general feeling of trust amongst people in a certain country. Redding and Venables (2004:68) claim to take their control variables from Knack and Keefer (1997) among others, but do not make use of their social capital indicators in their regressions. Including these alternative indicators gives excellent results. Significance of the socio-institutional controls increases sharply and R-squares are between 14 and 22 percent higher (see Table 5.4). Incorporating the Index of Economic Freedom and trust, shows that the latter are strongly significant and makes the variables "socialist rule" and the war dummy be automatically dropped from the regression.[8]

The main shortcoming in using the variable trust is the narrow sample size, which makes it difficult to compare with the other specifications counting with a broader sample of countries. This could lead to sample selection bias. The important point is that the highest adjusted r-square is obtained when trust in included in the regression. Then, the socialist and war dummies are automatically dropped. This shows that the concept of social capital, in this case in the form of trust, should not be neglected. Interestingly enough, the foreign market access coefficient is reversed. So the sign of the coefficient is also sensitive to the selection of socio-institutional variables. However,

[7] Civic engagement, another measure of social capital used in previous chapters, is not included in the regressions because it did not show statistical significance in a consistent way.
[8] The same is true for total market access (including domestic market as well as foreign).

we need to resort to other alternative socio-institutional control variables in order to explore the significance of the Market Access indicators deeper because of the short sample of the social capital indicator in this case. Unfortunately, the social capital indicator variable, trust from the World Value Surveys, does not offer the possibility to be explored in much more detail in a reasonable way despite its significance and high explanatory power due to the shortness in sample size, which would imply running out of degrees of freedom for any simple further manipulation of the variable. Nevertheless, we will come back to the usefulness of social capital related variables in the analysis further down in the chapter.

Another alternative (or, rather, complementary) exploration is to turn the attention to the widely used Governance Matters indicators gathered for recent years by the World Bank. The World Bank's Governance Matters indicators include: voice and accountability, political stability, government effectiveness, regulatory quality, rule of law, and control of corruption[9]. The effects of individual Governance Matters variables on total market access (which incorporates the domestic market as well as foreign) are presented in Table 5.5. Including the Governance Matters indicators one by one in the regressions produces switching signs of total market too. The sign of access to markets alternates depending on the choice of institutional variable. For voice and accountability, regulatory quality, rule of law, and control of corruption, the effect of access to markets into GDP is positive. For political stability and government effectiveness (the most statistically significant by far when all are included simultaneously), the effect of access to markets into GDP is negative[10]. When all of these variables are included simultaneously in the regression, substituting the poor performing "socialist rule" and war dummy, the sign of foreign and total market access is positive (see specification 7 in Table 5.5). It is worth noticing that only one of the Governance Matters variables remains statistically significant throughout: government effectiveness. This is why the effect of this variable as institutional control is analysed in a separate regression (specification 9 in Table 5.4). Including the variable government effectiveness from the Governance Matters V database reverses again the

[9] Data and detailed explanation of each variable can be obtained from the World Bank (Kaufmann, Kraay and Mastruzzi, 2006).

[10] Table 5.5b in the appendix repeats all the regressions in Table 5.5 using the original NBER-UN data instead. Once more, the sign of the market access coefficient switches from positive to negative in two occasions (both when we choose "voice" and when we choose "political stability" as controls) and from negative to positive when we include all the Governance Matters indicators simultaneously in the regression.

sign of the foreign market access coefficient, leaving it still within non-significance. Government effectiveness turns out to be strongly significant and produces an only slightly positive foreign market access coefficient. Intriguingly enough, if we run the same regression with total market access instead of foreign market access the sign of market access turns negative[11]. The problem of working with total instead of foreign market access is that the number of observations is reduced drastically to almost only half the sample size. This is why foreign market access is the preferred market access indicator. Furthermore, we have shown that the sign of access to markets alternates depending on the choice of institutional variable. So, the effect of the Market Access indicator is again unclear and we have learnt that alternative socio-institutional variables should not be ignored and deserve deeper exploration.

Overall, the evidence shows that Redding and Venables' results about the impact of the market access indicator are not always robust. The instrumental variables approach brings statistical significance to Market Access and the magnitude of the coefficient is similar to Redding and Venables (2004:69). However, the use of more standard institutional quality measures does not help to replicate their results, nor does the original NBER database. The regressions above show how sensitive the direction of the effect of access to markets into GDP is. Then, what is the real sign of the effect of access to markets? Given the impact and relevance of the paper by Redding and Venables (2004) introducing the market access indicators, it seems worth exploring what is driving the sensitivity of the results.[12] We investigate the sensitivity of the results with the aid of partition regressions on the government effectiveness, which exhibits both statistical significance and wide-enough sample of countries. The next sections of the chapter cover this sensitivity exploration.

5. X. WHAT LEADS THE RESULTS? PARTITION REGRESSIONS

Why are the Redding and Venables (2004) results so sensitive? What is the real effect of access to markets? This chapter has shown how standard socio-institutional controls produce switching signs of the market access indicator, which turns out to be not so

[11] Table 5.4b in the appendix reproduces the same results as in Table 5.4, –except regression 1 in Table 5.4, which corresponds to Redding and Venables' published results–, but this time using the original NBER-UN data. It produces similar results on the market access coefficient except for the regression incorporating the trust variable, which switches the sign of the market access coefficient once more. This may be due to the small sample available for this particular specification.

[12] According to Google Scholar, Redding and Venables (2004) paper had 410 citations as of 26/09/2008.

robust. In other words, the Redding and Venables result on the relevance of the market access indicator is not unconditional. The effect of the Governance Matters indicators – in particular, of the most significant government effectiveness–, deserves a deeper exploration. In this section, the paper explores the varying degrees of intensity and direction of the effect of government effectiveness on foreign market access.

Government effectiveness, taken from the Governance Matters V indicators, encloses some valuable information to be added to the analysis of access to markets. Limao and Venables (2001) point out that infrastructure is an important element of the costs of transportation, and it takes special relevance in the inland African countries. Their analysis of bilateral trade flows in African countries reveals that their poor volumes of trade are due to poor infrastructures (Limao and Venables, 2001:467). Other authors had already focused on the infrastructural component of high transportation costs and pointed out at inappropriate policies (Amjadi and Yeats, 1995). Inappropriate transport policies or the lack of effectiveness in implementing them cause poorer infrastructures for facilitating transport and trade. So, it seems reasonable to suggest that government effectiveness can play a role in explaining trade volumes.

Government effectiveness is the institutional variable that shows more robustness of all Governance Matters indicators. Recall, it is the only one that remains statistically significant when simultaneously regressing on all 6 of them. Overall, it has a positive effect on GDP per capita. Surprisingly, it produces a negative market access coefficient in both cases, foreign and total market access. A negative market access coefficient means that the more accessibility to markets a country has the poorer it is. This is the contrary of what one would have expected (Redding and Venables 2004 obtain a positive association between access to markets and GDP using alternative institutional variables); and, furthermore, it is counterintuitive. Therefore, it seems interesting to explore what is driving this result. Recall that other variables like political stability, risk of expropriation, or Index of Economic Freedom together with the variable "trust" also return negative market access coefficients. So, it is not an isolated phenomenon. This section focuses on the effect of one representative institutional variable (government effectiveness) into foreign market access.

Now, the question is the following: Is there an institutional quality threshold that countries need to reach before benefiting from good access to foreign markets?

There are a total of 181 observations for government effectiveness. The total sample of countries has been divided into several sub-samples according to their government effectiveness. Then, regressions (1) and (9) in Table 5.4 are run for each one of these sub-samples, corresponding to Redding and Venables' preferred specification with property rights protection, socialist and war dummies as institutional variables; and the author's preferred specification with property rights and government effectiveness, respectively. For extra robustness checks, both foreign and total market access have been tested, and both for NBER and CEP databases. A summary of results is presented in Table 5.6.

Table 5.6 displays the significance of the results using Redding and Venables' preferred institutional controls (property rights protection, socialist past, and recent war dummy) as in Table 5.4, equation (1)[13]. It derives from the table that the upper quartile of the observations enjoys the most significant effect of geographical access to markets, while the lower 75 percent of the sample never gets a significant coefficient when analysed separately. Therefore, the whole sample of countries has been partitioned into 2 groups: those scoring highest in the government effectiveness indicator (upper quartile of the observations), which corresponds to .50 or above in the government effectiveness index, ranging from -2.5 to 2.5 approximately (Kaufman, Kraay, and Mastruzzi, 2006); and the rest (lower 3 quartiles). The division between the two groups of observations according to the government effectiveness ranking has been made effective through the creation of a dummy variable operating in the upper quartile of the sample. Then, an interaction term between market access and the government effectiveness dummy has been introduced in order to account for the distinctive nature of the upper sub-sample. This interaction term operates as a slope dummy on the market access coefficient.

Table 5.7 shows the baseline regressions, (1) and (4), augmented with the market access-government effectiveness interaction term, in (2), (3), (5) and (6)[14] . Notice that foreign market access has a small and non-significant coefficient, while the interaction

[13] If I use my preferred institutional variables (property rights and government effectiveness), as in Table 5.4, equation (6), I get non-significance throughout. This means that the market access coefficient is not significant, except for Redding and Venables' choice of institutional variables.

[14] Running the regressions with Foreign Market Access calculated with the NBER-UN data instead of with the CEP data adds a few more observations to the effective sample. Regressions have also been run for total market access (NBER-UN data) and results are again similar.

term (lfmaCefhighest) enjoys a positive and highly significant coefficient of the magnitude of 0.05 to 0.08. Recall the interaction term is the result of multiplying the foreign market access index by the government effectiveness dummy operating in the upper quartile of the observations, and should be interpreted as extra slope for those observations within the upper 25 percent sub-sample. The same results are obtained when market access and the interaction term are instrumented with the distance to the main world markets (see specifications (3) and (6). In this case, the market access coefficient is positive but once more non-significant, but its interaction term with government effectiveness is taking all the significance with the magnitudes of the coefficient having even risen to 0.10. This result holds even if the institutional variables are represented by trust only, which reduces the number of observations to less than half the sample (see specification (6) in Table 5.7). Therefore, most of the effect captured by foreign market access overall is due to those observations concentrated on the upper government effectiveness quartile. So, institutional quality matters and there seems to be a threshold below which a country is not able to effectively enjoy the benefits of good location.

5. XI. EMPIRICAL COUNTERFACTUALS: CHANGES IN COUNTRY CHARACTERISTICS

We know from last section that if one would perform the hypothetical experiment of moving Zimbabwe to Europe its potential 80 percent rise in GDP per capita would not be effective, because Zimbabwe is not a high government effectiveness country.

In this section, Market Access has been computed based on specific country characteristics instead of country dummies, so that one can evaluate the effect of a specific characteristic (for instance, what is the isolated effect of distance to markets?). I did this for foreign and domestic market access, foreign and domestic supplier access, and all of them using the two alternative datasets UN and CEP (the latter shown in tables, for consistency with Redding and Venables, 2004). The table is based on the preferred control variables specified by Redding and Venables, with the intention of getting comparable results[15].

[15] Redding and Venables (2004:76) demonstrated that the use of alternative parameters for intermediate goods share (α) and elasticity of substitution (σ) produces very small changes in predictions.

The figures in Table 5.8 display the predicted percentage change in levels of GDP *per capita* when one does the experiment of hypothetically moving a given country to central Europe. Four countries have been selected for this experiment: Zaire, Sri Lanka, Zimbabwe, and New Zealand; ordered according to their government effectiveness indicator. Zaire and New Zealand represent one very low and one very high government effectiveness observations, respectively. The 2 middle-ranged countries are case studies by Redding and Venables (2004). Regressions using country characteristics where used to compute the changes in predicted GDP, which give a slightly negative (but not significant) coefficient to foreign market access[16]. This is why the GDP changes with no slope dummy are slightly negative. Columns (1) and (2) show the potential rise in GDP, as calculated by Redding and Venables (2004: Table 5.7). This methodology does not make a difference between good and bad institutions' countries. Potential increases in GDP are very high, often above 100 percent (i.e. income would double) with respect to predicted GDP with actual distances and common borders. Columns (3) and (4) take into account institutional handicaps. Poor and middle-range government effectiveness countries below the 75 percent threshold practically do not change their GDP even after endowing them with European borders and distances. Slightly negative numbers are due to a technicality, —the slightly negative though not significant coefficient of market access in the regressions by country characteristics (not shown)—, but the idea is that it reflects a practically null change. On the contrary, should New Zealand enjoy European distances to markets, the predictions say that it would enjoy an increase in GDP of around 20 percent, which seems more reasonable. This would be once we adjust for the effect of good institutional health. Columns (4) and (6) show the predicted percentage change in GDP per capita after reversing the status of every country's high government effectiveness dummy. This is, countries belonging to the low and middle government effectiveness scores are given high quality institutional dummy status, and countries with high government effectiveness are deprived from it. The overall result is that when countries with low and middle government effectiveness are artificially assigned to the high value group, their GDP doubles as when we made no institutional quality distinction. So, saying that if we would move a distant country to central Europe the GDP would almost double is an exaggeration and cannot be generalised. Redding and Venables's results have been generically validated but should have taken into account the big difference that the detrimental effect of low quality institutions makes. Even if

[16] The set of estimations with specific country characteristics is not displayed, but can be made available upon request.

we would move Zimbabwe to central Europe (as it is) its GDP would not increase by 80 percent because it does not reach the minimum institutional quality threshold. Institutional quality and, in particular, government effectiveness, makes a big difference. Among the 4 countries studied only New Zealand belongs to the upper government effectiveness quartile. This is reflected in the drastic fall in GDP (above 88 percent fall) when its privileged institutional quality is experimentally removed[17].

Overall, these four country cases illustrate how the possibility of a potential rise in GDP per capita when moving closer to markets being actually realised is very sensitive to institutional quality (in particular, to government effectiveness).

5. XII. CONCLUSIONS

This chapter provides an example of how and why social capital should not be left out of empirical economic theory. The inclusion of social capital proxies in the regressions improves the fit and opens up an alternative way of thinking about how society and the informal institutional setup matter for economic performance. Given the non-robustness of the results for the market access indicator in the mid-1990s, the Social Development Index (SDI) facilitates testing the theory of the relevance of access to markets for a different benchmark year. Testing the market-access-matters hypothesis for the mid-1960s, for which the SDI is available, gives the opportunity to check whether the results for the 1990s can be transferred to the Golden Age. The results show that the regression analysis gives very similar results for both periods.

The direct effect of access to markets on GDP per capita is less robust than what we initially thought. The original Redding and Venables (2004) results cannot be replicated with the NBER database, nor can they with alternative control institutional measures. The income regressions produce a switching sign of the coefficient depending on which institutional variables does one choose as controls. Repeating the analysis for an alternative benchmark year (1965) does not lead to more conclusive results about the significance of the effect of the market access indicator. Still, it reveals weakly positive

[17] Changes in distances and borders have been performed both from the exporter and the importer perspectives. If changes as an exporter only are considered, the magnitudes of the change in GDP per capita are much more moderate. For instance, Table 5.5, column (2), Zimbabwe would be 34.73 and New Zealand 45.52. Nevertheless, I understand changing a given country characteristics should be performed at both ends: both from the exporter and the importers' perspective.

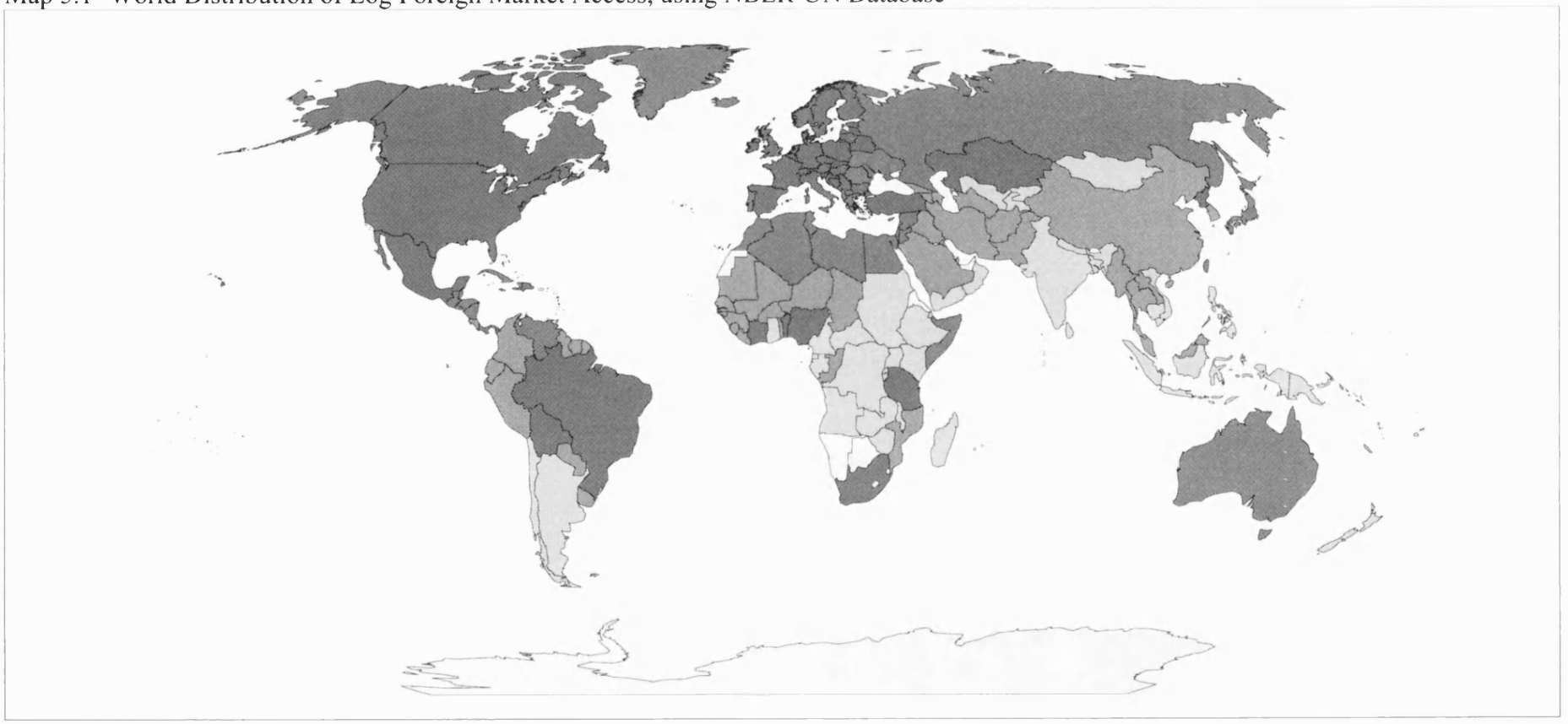
coefficient in all income regressions, which suggests a positive association between market access and GDP per capita at that time.

However, results come much more strongly when we look at geographical bilateral distances to the main World markets, which seem to act indirectly through the Market Access indicators on real GDP per capita differentials. The strongly positive coefficients of access to markets when instrumented by distance to the main world markets suggest that the Market Access indicators have a positive effect on economic growth indirectly through distance.

The results show that access to markets mattered as much for the 1960s as it did for the 1990s. This means that during times of increasing international trade volumes such as the Golden Age, the good geographic pre-disposition of the countries in terms of distance did matter for getting richer through international trade. It is well-known that commercial interactions were more intense within commercial geographical blocks (for instance, within the European Economic Community), and pure geographic characteristics (which is what Market Access captures) do seem to have been a driving force behind the increase in per capita incomes at that time. Of course, other factors such as preferential commercial agreements might have played an equally important role.

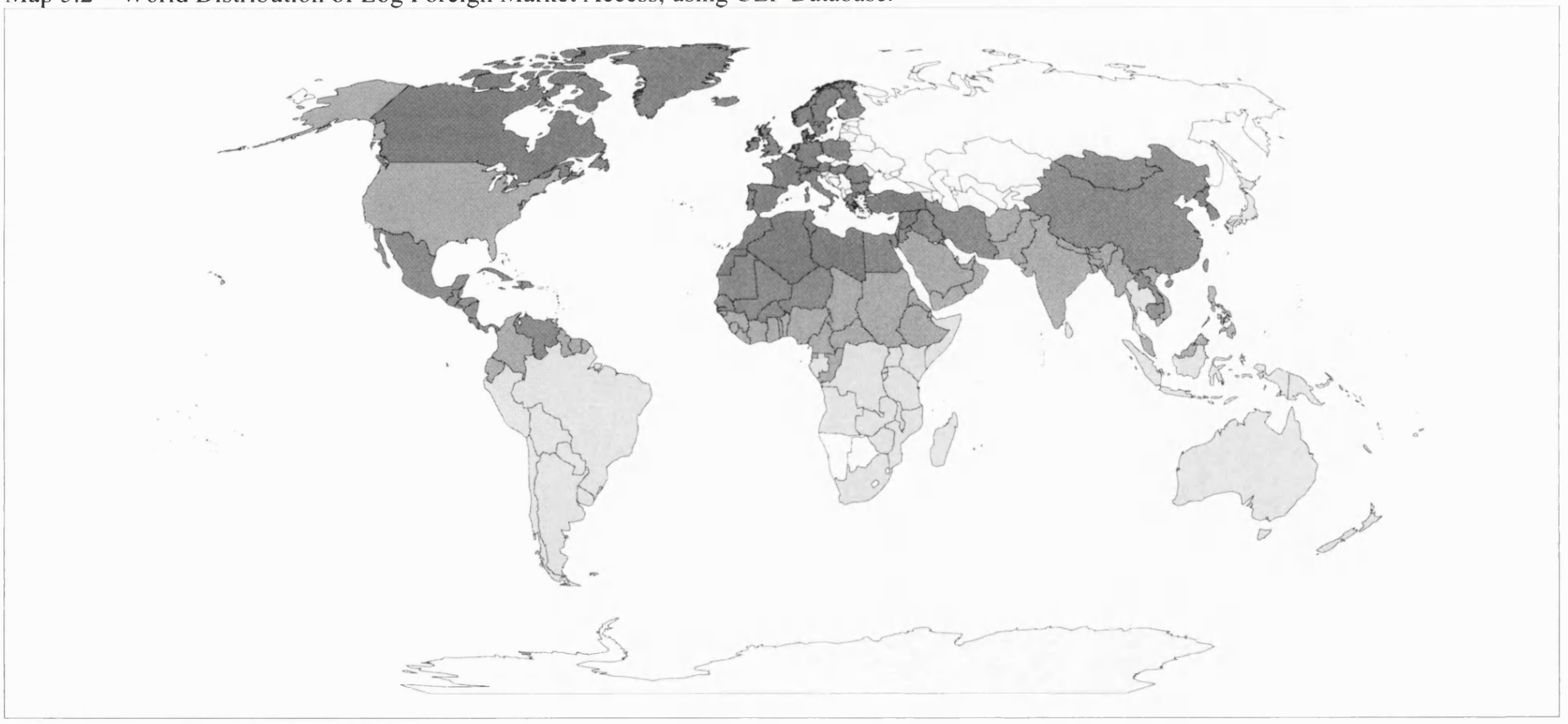
Having a closer look at the Market Access indicators, partition regressions show that the benefits of geographical access to markets operate with varying degrees of intensity, and sometimes negatively. So, the Redding and Venables (2004) results cannot be generalised. The institutional setup of a country is not neutral, and can operate as to enhance or deprive a country from the benefits of a good geographical positioning. Countries with low government effectiveness do not show an ability to benefit from a good geographical location, while countries with high government effectiveness (above the .50 threshold on a -2.5 to 2.5 scale) are more sensitive to good access to markets.

Map 5.1- World Distribution of Log Foreign Market Access, using NBER-UN Database



Note: Equal count criterion in choosing the range.

Map 5.2 – World Distribution of Log Foreign Market Access, using CEP Database.



Countries by lfmaCEP

- 13.110.000 to 15.380.000 (43)
- 12.710.000 to 13.110.000 (35)
- 12.470.000 to 12.710.000 (41)
- 11.670.000 to 12.470.000 (42)

Note: Equal count criterion in choosing the range.

Table 5.1 - Real GDP per capita, Market Access (Foreign and Total), Economic Geography, and Social Development

Log GDP pc	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Observations	57	57	59	59	59	59	65	65
Year	1964	1964	1965	1965	1965	1965	1965	1965
Log FMA1962	.3000	-	-	-	.0770	-	-	-
	(.2400)				(.1113)			
Log TMA1962	-	.2995	-	-	-	.0773	-	-
		(.2400)				(.1113)		
Log FMA1964	-	-	.0864	-	-	-	-	-
			(.1081)					
Log TMA1964	-	-	-	.0867	-	-	-	-
				(.1081)				
Log FMA1965	-	-	-	-	-	-	.0009	-
							(.0013)	
Log TMA1965	-	-	-	-	-	-	-	.0009
								(.0013)
Log Hydroc pc	-.0618*	-.0618*	.0262	.0261	.0265	.0264	.0256*	.0256*
	(.0355)	(.0355)	(.0156)	(.0160)	(.0160)	(.0160)	(.0147)	(.0147)
Arable land	-.1066	-.1064	.1193**	.1193 **	.1178 **	.1178**	.1042*	.1042*
	(.1071)	(.1071)	(.0522)	(.0521)	(.0534)	(.0533)	(.0524)	(.0524)
N minerals	-.0055	-.0055	-.0011	-.0011	-.0011	-.0011	-.0032	-.0032
	(.0221)	(.0221)	(.0094)	(.0094)	(.0094)	(.0094)	(.0099)	(.0099)
% tropics	.2597	.2599	.0139	.0139	.0066	.0066	-.0007	-.0008
	(.3339)	(.3339)	(.1389)	(.1389)	(.1409)	(.1408)	(.1264)	(.1264)
Malaria 1966	-.2039	-.2042	-.1087	-.1087	-.1071	-.1070	-.0655	-.0650
	(.3801)	(.3802)	(.1720)	(.1720)	(.1736)	(.1736)	(.1552)	(.1553)
Social dev	-.0934	-.0934	.5222***	.5221***	.5228***	.5228***	.5160***	.5160***
	(.1720)	(.1720)	(.0624)	(.0624)	(.0623)	(.0623)	(.0644)	(.0644)
Estimation	OLS	OLS	OLS	OLS	OLS	OLS	OLS	OLS
R^2	.1283	.1282	.6731	.6731	.6719	.6719	.6347	.6348
F(.)	.97	.97	18.75	18.76	18.91	18.91	15.27	15.27
Prob>F	.4642	.4650	.0000	.0000	.0000	.0000	0.0000	0.0000

Notes: Constant not shown. Heteroskedasticity corrected White-robust standard errors. Regressions (1) and (2) with 1964 GDP do not pass the joint F-test. This might be due to simultaneity bias. Therefore, it is preferable to run the regressions on 1965 GDP levels, which pose no statistical problem. *Statistically significant at the 10% level. ***Statistically significant at the 1% level

Table 5.2 – 1996 Real GDP per capita, Distance to Main World Markets, and Proxy for Social Capital (Trust)

Log GDP pc	(1)	(2)	(3)	(4)
Observations	40	40	40	40
Year	1996	1996	1996	1996
LogFMA1994	.3016**	-	-	-
	(.1408)			
LogTMA1994	-	.3021**	-	-
		(.1410)		
LogFMA1996	-	-	.2896**	-
			(.1341)	
LogTMA1996	-	-	-	.2899**
				(.1342)
Loghydroc pc	.0027	.0027	.0031	.0031
	(.0193)	(.0194)	(.0194)	(.0194)
Arable Land	-.0139	-.0137	-.0139	-.0137
	(.0709)	(.0709)	(.0701)	(.0701)
N minerals	.0027	.0027	.0035	.0034
	(.0133)	(.0133)	(.0133)	(.0133)
% Tropics	-.5968	-.5963	-.6000	-.5997
	(.3671)	(.3671)	(.3592)	(.3591)
Malaria 1994	-1.640***	-1.640***	-1.606***	-1.606***
	(.4393)	(.4392)	(.4296)	(.4295)
Trust	.0171**	.0171**	.0171***	.0171***
	(.0064)	(.0064)	(.0063)	(.0063)
Estimation	2SLS	2SLS	2SLS	2SLS
R^2	0.7509	0.7510	0.7538	0.7539
F(.)	0.0000	0.0000	0.0000	0.0000
Prob>F	20.13	20.11	20.22	20.22

Constant not shown. Heteroskedasticity White-robust standard errors. Instruments are the Log of distance capital to capital of each country to the USA, Belgium, Japan and the minimum of the 3. ***Statistically significant at the 1% level.

Table 5.3 – 1965 Real GDP per capita, Distance to Main World Markets, and Proxy for Social Capital (Soc Dev)

Log GDP pc	(1)	(2)	(3)	(4)
Observations	59	59	59	59
Year	1965	1965	1965	1965
LogFMA1962	.3283**	-	-	-
	(.1623)			
LogTMA1962	-	.3287**	-	-
		(.1622)		
LogFMA1965	-	-	.3052**	-
			(.1413)	
LogTMA1965	-	-	-	.3054**
				(.1412)
Loghydroc pc	.0192	.0192	.0203	.0203
	(.0174)	(.0174)	(.0170)	(.0170)
Arable Land	.1046*	.1047*	.1079**	.1080**
	(.0547)	(.0547)	(.0523)	(.0523)
N minerals	-.0002	-.0002	.0004	.0004
	(.0101)	(.0101)	(.0100)	(.0100)
% Tropics	-.0237	-.0236	-.0071	-.0070
	(.1489)	(.1489)	(.1472)	(.1472)
Malaria 1966	-.0497	-.0498	-.0530	-.0531
	(.1979)	(.1979)	(.1952)	(.1952)
Soc Dev	.5165***	.5163***	.5094***	.5093***
	(.0675)	(.0675)	(.0669)	(.0669)
Estimation	2SLS	2SLS	2SLS	2SLS
R^2	0.6328	0.6328	0.6425	0.6425
F(.)	0.0000	0.0000	0.0000	0.0000
Prob>F	17.05	17.05	17.28	17.28

Constant not shown. Heteroskedasticity White-robust standard errors. Instruments are the Log of distance capital to capital of each country to the USA, Belgium, Japan and the minimum of the 3. *Statistically significant at the 10% level. **Statistically significant at the 5% level. ***Statistically significant at the 1% level.

Table 5.4 – Log Current Price GDP per capita 1996, and Foreign Market Access

LogGDPpc	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Observations	91	91	96	58	91	96	25	40	25
Year	1996	1996	1996	1996	1996	1996	1996	1996	1996
LogFMACEP	.215** (.063)	.1281(.0835)	-	-.0200(.1655)	.0985 (.0683)	-	-.0023(.0570)	-	.0063 (.0703)
LogFMA	-	-	.2433**	-	-	.1939*	-	.3016**	-
			(.1090)			(.1011)		(.1408)	
Loghydroc pc	.019	.0402***	.0387 **	.0405**	.0543***	.0515***	.0042	.0027	.0023
	(.015)	(.0150)	(.0148)	(.0193)	(.0142)	(.0143)	(.0125)	(.0193)	(.01456)
Arable Land	-.050	-.0462	-.0691	-.0870	-.0446	-.0632	-.0581	-.0139	-.0548
	(.066)	(.0500)	(.0471)	(.0705)	(.0432)	(.0431)	(.0348)	(.0709)	(.0409)
N minerals	.016**	.0051	.0045	.0053	.0061	.0051 .0106	.0019	.0027	.0021
	(.008)	(.0108)	(.0104)	(.0122)	(.0109)		(.0120)	(.0133)	(.0122)
% tropics	-.057	-.1901	-.1855	-.4083**	-.3150	-.3272*	.1375	-.5968	.1293
	(.239)	(.2131)	(.2018)	(.1870)	(.2017)	(.1944)	(.2954)	(.3671)	(.3168)
Malaria 1994	-1.107**	-1.159***	-1.141***	-1.315***	-.7670***	-.7898***	-2.152***	-1.640***	-2.185***
	(.282)	(.2135)	(.2163)	(.2401)	(.2226)	(.2157)	(.2555)	(.4393)	(.2745)
Property right	-.445**	-.4517***	-.4243***	-	-	-	-	-	-
	(.091)	(.0824)	(.0809)						
Risk expropr	-	-	-	.2078***	-	-	-	-	-
				(.0511)					
Econ Free	-	-	-	-	-.8216***	-.7579***	-.7680***	-	-.8146**
					(.1387)	(.1410)	(.1690)		(.2890)
Socialist rule	-.210(.191)	-.0995(.1739)	-.1472(.1877)	-.0419(.2690)	.0097 (.1980)	-.0680(.2095)	(dropped)	-	(dropped)
War dummy	-.052 (.169)	-.0183(.1614)	-.0022(.1603)	-.0336(.1485)	-.0768(.1737)	-.0823(.1744)	(dropped)	-	(dropped)
Trust	-	-	-	-	-	-	.0171***	.0171**	.0187*
							(.0058)	(.0064)	(.0095)
Gvt Effective	-	-	-	-	-	-	-	-	-.0435(.2231)
Estimation	OLS	OLS	2SLS	OLS	OLS	2SLS	OLS	2SLS	OLS
R^2	.766	.7702	.7830	.7119	.7920	.7971	.9298	.7509	.9301
F(.)	47.77	40.62	54.48	20.79	42.11	51.48	76.59	20.13	65.39
Prob>F	.000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000

Notes: Constant not shown. Heteroskedasticity corrected White-robust standard errors in parentheses. Due to small sample properties, standard errors in regressions (5) and (6) have been double-checked with bootstrapping (200 replicas), giving the same significance results for regression (5); only the Index of Economic Freedom remains significant for regression (6). *Statistically significant at the 10% level. **Statistically significant at the 5% level. ***Statistically significant at the 1% level.

Table 5.5 – Log of current price real GDP pc 1996 (PWT 6.1), Market Access (CEP), and Governance Matters V

Log GDP pc	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Observations	45	45	45	45	45	44	44
Year	1996	1996	1996	1996	1996	1996	1996
LogTMACEP	.0204	-.0066	-.0131	.0539	.0068	.0186	.0944
	(.1138)	(.1198)	(.1208)	(.1195)	(.1200)	(.1377)	(.1209)
Loghydroc pc	.0829***	.0804***	.0750***	.0781***	.0750***	.0722***	.0631**
	(.0231)	(.0236)	(.0220)	(.0232)	(.0247)	(.0233)	(.0291)
Arable Land	-.0848	-.1229	-.0830	-.0621	-.0866	-.0483	.0565
	(.1013)	(.1068)	(.1154)	(.1108)	(.1094)	(.1433)	(.1238)
Nminerals	-.0088	.0006	-.0034	-.0061	-.0066	.0007	-.0029
	(.0181)	(.0180)	(.0162)	(.0200)	(.0191)	(.0176)	(.0186)
% Tropics	-.2743	-.3233	-.0488	-.2969	-.3033	-.1221	.0716
	(.2552)	(.2855)	(.2778)	(.2797)	(.2675)	(.2728)	(.3211)
Malaria 1994	-.9439***	-.9820***	-.9677***	-.8030***	-1.007***	-1.067***	-1.052***
	(.2448)	(.2628)	(.2493)	(.2676)	(.2524)	(.2704)	(.3140)
Property right	-.3603***	-.3858***	-.2498*	-.2979**	-.4557***	-.3168**	-.1023
	(.1182)	(.0953)	(.1243)	(.1285)	(.1194)	(.1302)	(.1508)
Voice	.2594***	-	-	-	-	-	.1192
	(.0901)						(.2972)
Polit. stability	-	.1890**	-	-	-	-	.0129
		(.0856)					(.1279)
Gvt. Effective	-	-	.3955***	-	-	-	.5216**
			(.1401)				(.2124)
Regulatory	-	-	-	.3569*	-	-	.1920
				(.1771)			(.2575)
Rule of Law	-	-	-	-	.1036	-	-.2146
					(.1253)		(.1750)
Contrl corrup	-	-	-	-	-	.2496**	-.1082
						(.1193)	(.2393)
Estimation	OLS	OLS	OLS	OLS	OLS	OLS	OLS
R^2	.8285	.8253	.8433	.8299	.8145	.8253	.8599
F(.)	37.41	35.98	48.94	40.28	30.53	37.76	44.00
Prob>F	.0000	.0000	.0000	.0000	.0000	.0000	.0000

Constant not shown. Heteroskedasticity corrected White-robust standard errors in parentheses. *Statistically significant at the 90% level. **Statistically significant at the 95% level. ***Statistically significant at the 99% level.

Table 5.6 – Exploring the government effectiveness threshold with several sub-samples:
Is Market Access significant?

SAMPLE	10%	5%	1%	also CEP
19/163 (all but upper and lower 10%)	×	×	×	No
1/45 (lower 25%)	×	×	×	No
1/135 (lower 75%)	×	×	×	No
46/181 (upper 75%)	✓	✓	×	No
1/90 (lower 50%)	✓	✓	×	No
91/181 (upper 50%)	✓	✓	×	No
136/181 (upper 25%)	✓	✓	✓	Occasionally
1/162 (all but upper 10%)	×	×	×	No

There is a total of 181 observations for government effectiveness. Columns represent 10, 5, and 1 percent significant levels respectively. Results hold for both foreign and total market access. Last column means "Also using the CEP database?", which has a narrower sample. "Occasionally" means "Yes" for Foreign Market Access at 10% only.

Table 5.7 – Log of current price real GDP pc 1996 (PWT 6.1), Foreign Market Access (CEP), and partitioned government effectiveness

Log GDP pc	(1)	(2)	(3)	(4)	(5)	(6)
Observations	92	92	92	92	92	38
Year	1996	1996	1996	1996	1996	1996
logFMA CEP	.1323	.0337	.0311	.0114	.0035	.0288
	(.0839)	(.0663)	(.0736)	(.0666)	(.0639)	(.1006)
LogFMA CEP effechighest	-	.0795***	.1052***	-	.0505***	.0951**
		(.0104)	(.0350)		(.0146)	(.0415)
Log hydroc pc	.0415***	.0543***	.0586***	.0529***	.0561***	.0501*
	(.0150)	(.0126)	(.0143)	(.0128)	(.0118)	(.0249)
Arable Land	-.0446	-.0481	-.0465	-.0481	-.0483	-.0400
	(.0502)	(.0388)	(.0393)	(.0432)	(.0399)	(.0703)
N minerals	.0060	-.0063	-.0104	.0060	-.0015	-.0063
	(.0107)	(.0087)	(.0112)	(.0083)	(.0086)	(.0156)
% Tropics	-.1802	.1499	.2795	.1291	.1963	.1745
	(.2113)	(.1614)	(.2217)	(.1645)	(.1485)	(.5357)
Malaria 1994	-1.184***	-1.240***	-1.254***	-1.065***	-1.144***	-1.623***
	(.2072)	(.1758)	(.1713)	(.1793)	(.1641)	(.2686)
Property rights	-.4490***	-.2253***	-.1509	-.1472*	-.1376*	-
	(.0823)	(.0773)	(.1219)	(.0828)	(.0795)	
Socialist	-.0984	.0077	.0333	-	-	-
	(.1748)	(.1761)	(.1900)			
War dummy	-.0126	-.0768	-.0952	-	-	-
	(.1599)	(.1187)	(.1213)			
Trust	-	-	-	-	-	.0073(.0070)
Gvt Effective	-	-	-	.5322***	.2996***	-
				(.0837)	(.1108)	
Estimation	OLS	OLS	2SLS	OLS	OLS	2SLS
R^2	.7753	.8487	.8407	.8437	.8605	.7954
F(.)	42.41	61.26	47.60	59.32	63.26	25.26
Prob>F	.0000	.0000	.0000	.0000	.0000	.0000

Constant not shown. Heteroskedasticity corrected White-robust standard errors in parentheses. In specifications (3) and (6), Log FMA CEP and Log FMA CEP effechighest have been instrumented with the log distance to USA, Belgium, Japan, and the minimum of the 3 in a first stage regression. *Statistically significant at the 90% level. **Statistically significant at the 95% level. ***Statistically significant at the 99% level.

Table 5.8 - Percentage changes in predicted *levels* of GDP per capita 1996

countries, from less to more government effectiveness	(1) Distance only of Central Europe	(2) Distance &common borders of central Europe	(3) Distance only of Central Europe), Instit. threshold considered	(4) Distance only to Central Europe, Instit. threshold considered → Instit. Change	(5) Distance &common borders of central Europe, Instit. threshold considered	(6) Distance &common borders of central Europe, Instit. thresthold considered → Instit. Change
Zaire	81.59	100.51	-2.80	105.35	-2.50	103.89
Sri Lanka	103.33	132.63	-3.30	107.72	-3.39	108.15
Zimbabwe	96.91	108.96	-3.28	104.94	-2.82	102.73
New Zealand	131.22	146.02	20.83	-88.67	19.42	-88.37

Notes: Specification as suggested by Redding and Venables 2004 (Table 5.5, column 3), including Foreign Market and Supplier Access and control variables, with no constant and parameters constraint as in their equation (21), α =0.5, σ=10. Market and Supplier access calculated using country characteristics instead of country dummies in order to allow for specific feature effects. Calculations based on CEP database. Distance to central and Central Europe consist of imputing the distances (and common borders when specified) corresponding to Hungary. The institutional change in columns (4) and (5) consists of giving value 1 when the government effectiveness dummy is 0, and giving it value 0 when it is 1 (the latter only for New Zealand).

APPENDIX 5.A
LIST OF VARIABLES IN THE TABLES

Self constructed

lfma ln of foreign market access 1994, constructed from the NBER-UN database on bilateral trade and the STATA code from Redding and Venables (2004)[18].

lma_c ln of total market access 1994, calculated as the logarithm of the sum of domestic plus foreign market access[19].

lfmaCEP ln of foreign market access using the CEP cleaned up version of the NBER-UN database

lma_cCEP ln of total market access using the CEP cleaned up version of the NBER-UN database.

lfmaccCEP ln of foreign market access calculated using country characteristics instead of country dummies (CEP database).

lfsaccCEP ln of foreign supply access calculated using country characteristics instead of country dummies (CEP database).

Partition variables: dummy and interaction term

Instrumental variables: minimum of the 3 log distances to the USA, Belgium and Japan

From data sources

Endogenous and control variables:

lcgdp ln of current price real GDP per capita in 1996, from the Penn World Tables 6.1

land Arable land area per capita (ln is used in the regressions)

lhcpc Hydrocarbons per capita (ln is used in the regressions)

nminerals Number of minerals

tropicar Fraction of land in the geographical tropics

malfal Prevalence of malaria

re Risk of expropriation index. 0 to 10 scale (real numbers).

---

[19] For domestic market access, MA(3) as in Redding and Venables (2004) is used; this is, the outcome to the TOBIT specification which contemplates the truncated nature of the bilateral trade data.

pr Property rights 1996, from the Index of Economic Freedom. 1 to 5 scale (natural numbers).

free Score for Index of Economic Freedom 1996.

socialst Socialist rule during 1950-1995

wardum External war 1960-1985

voice voice and accountability 1996, from the World Bank "Governance Matters V" database.

polsta political stability 1996, from the World Bank "Governance Matters V" database.

effec government effectiveness 1996, from the World Bank "Governance Matters V" database.

reg regulatory quality 1996, from the World Bank "Governance Matters V" database.

RoL rule of law 1996, from the World Bank "Governance Matters V" database.

contrl control of corruption 1996, from the World Bank "Governance Matters V" database.

trustkk Trust (several years) as in Knack and Keefer (1997)

Instrumental variables:

ldistcapUSA Distance to the USA, capital city to capital city, from the *Centre d'Études Prospectives et d'Informations Internationales* (CEPII), (ln used in the regressions)

ldistcapBEL Distance to Belgium, capital city to capital city, from the *Centre d'Études Prospectives et d'Informations Internationales* (CEPII), (ln used in the regressions)

ldistcapJPN Distance to Japan, capital city to capital city, from the *Centre d'Études Prospectives et d'Informations Internationales* (CEPII), (ln used in the regressions)

APPENDIX 5.B

LIST OF COUNTRIES AND TERRITORIES

In alphabetical order according to the ISO3 World Bank classification code:

iso3	name of country/territory
ABW	ARUBA,
AFG	AFGHANISTAN
AGO	ANGOLA
AIA	ANGUILLA
ALB	ALBANIA
AND	ANDORRA
ANT	NETHERLANDS ANTILLES
ARE	UNITED ARAB E.
ARG	ARGENTINA
ARM	ARMENIA
ASM	AMERICAN SAMOA
ATF	FRENCH SOUTHERN TERRITORIES
ATG	ANTIGUA AND BARBUDA
AUS	AUSTRALIA
AUT	AUSTRIA
AZE	AZERBAIJAN
BDI	BURUNDI
BEL	BELGIUM
BEN	BENIN
BFA	BURKINA FASO
BGD	BANGLADESH
BGR	BULGARIA
BHR	BAHRAIN
BHS	BAHAMAS
BIH	BOSNIA AND HERZEGOVINA
BLR	BELARUS
BLX	BELGIUM (INCLUDES LUXEMBURG)
BLZ	BELIZE
BMU	BERMUDA
BOL	BOLIVIA
BRA	BRAZIL
BRB	BARBADOS
BRN	BRUNEI
BTN	BHUTAN
BWA	BOTSWANA
CAF	CENTRAL AFR.R.
CAN	CANADA
CCK	COCOS (KEELING) ISLANDS
CHA	CHANNEL ISLANDS
CHE	SWITZERLAND
CHL	CHILE
CHN	CHINA
CIV	IVORY COAST
CMR	CAMEROON
COG	CONGO
COK	COOK ISLANDS

COL COLOMBIA
COM COMOROS
CPV CAPE VERDE IS.
CRI COSTA RICA
CUB CUBA
CXR CHRISTMAS ISLAND
CYM CAYMAN ISLANDS
CYP CYPRUS
CZE CZECH REPUBLIC
CZS CZECHOSLOVAKIA
DDR GERMAN DEMOCRATIC REPUBLIC
DEU GERMANY, WEST
DJI DJIBOUTI
DKF DENMARK (INCLUDES FAEROE ISLANDS)
DMA DOMINICA
DNK DENMARK
DOM DOMINICAN REP.
DRG GERMANY, EAST
DZA ALGERIA
ECU ECUADOR
EGY EGYPT
ERI ERITREA
ESH WESTERN SAHARA
ESP SPAIN
EST ESTONIA
ETF ETHIOPIA (INCLUDES ERITREA)
ETH ETHIOPIA
FIN FINLAND
FJI FIJI
FLK FALKLAND ISLANDS (MALVINAS)
FRA FRANCE
FRO FAEROE ISLANDS
FSM MICRONESIA
GAB GABON
GBC GUINEA-BISSAU (INCLUDES CAPE VERDE)
GBR U.K.
GEO GEORGIA
GER GERMANY
GHA GHANA
GIB GIBRALTAR
GIN GUINEA
GLP GUADELOUPE
GMB GAMBIA
GNB GUINEA-BISS
GNQ EQUATORIAL GUINEA
GPM GUADELOUPE (INCLUDES MARTINIQUE)
GRC GREECE
GRD GRENADA
GRL GREENLAND
GTM GUATEMALA
GUF FRENCH GUIANA
GUM GUAM

GUY GUYANA
HKG HONG KONG
HND HONDURAS
HRV CROATIA
HTI HAITI
HUN HUNGARY
IDM INDONESIA (INCLUDES MACAU)
IDN INDONESIA
IND INDIA
IOM ISLE OF MAN
IOT BRITISH INDIAN OCEAN TERRITORY
IRL IRELAND
IRN IRAN
IRQ IRAQ
ISL ICELAND
ISR ISRAEL
ITA ITALY
JAM JAMAICA
JOR JORDAN
JPN JAPAN
KAZ KAZAKHSTAN
KEN KENYA
KGZ KYRGYZSTAN
KHM CAMBODIA
KIR KIRIBATI
KIZ KIRIBATI (INCLUDES SOLOMON ISLANDS, TONGA, TUVALU)
KNA ST.KITTS&NEVIS
KOR KOREA, REP.
KWT KUWAIT
LAO LAOS
LBN LEBANON
LBR LIBERIA
LBY LIBYA
LCA ST.LUCIA
LIE LIECHTENSTEIN
LKA SRI LANKA
LSO LESOTHO
LTU LITHUANIA
LUX LUXEMBOURG
LVA LATVIA
LWI KOSOVO
MAC MACAO
MAR MOROCCO
MAY MAYOTTE
MCO MONACO
MDA MOLDOVA
MDG MADAGASCAR
MDV MALDIVES
MEX MEXICO
MHL MARSHALL ISLANDS
MKD MACEDONIA
MLI MALI

MLT MALTA
MMR MYANMAR
MNG MONGOLIA
MNP NORTHERN MARIANA ISLANDS
MOZ MOZAMBIQUE
MRT MAURITANIA
MSR MONTSERRAT
MTQ MARTINIQUE
MUS MAURITIUS
MWI MALAWI
MYS MALAYSIA
NAM NAMIBIA
NCL NEW CALEDONIA
NCZ NEW CALEDONIA (INDLUDES FRENCH POLYNESIA, VANUATU)
NER NIGER
NFK NORFOLK ISLAND
NGA NIGERIA
NIC NICARAGUA
NIU NIUE
NLD NETHERLANDS
NOR NORWAY
NPL NEPAL
NRU NAURU
NZL NEW ZEALAND
OMN OMAN
PAK PAKISTAN
PAL PALESTINE
PAN PANAMA
PCI TRUST TERRITORY OF THE PACIFIC ISLANDS
PCN PITCAIRN
PER PERU
PHL PHILIPPINES
PLW PALAU
PNG PAPUA N.GUINEA
POL POLAND
PRI PUERTO RICO
PRK KOREA, DEM. REP.
PRT PORTUGAL
PRY PARAGUAY
PYF FRENCH POLYNESIA
QAT QATAR
REU REUNION
ROM ROMANIA
RUS U.S.S.R.
RWA RWANDA
SAU SAUDI ARABIA
SCG SERBIA AND MONTENEGRO
SDN SUDAN
SEN SENEGAL
SER SERBIA
SGP SINGAPORE
SHN SAINT HELENA

SLB SOLOMON IS.
SLE SIERRA LEONE
SLV EL SALVADOR
SMR SAN MARINO
SOM SOMALIA
SPM SAINT PIERRE AND MIQUELON
STP SAO TOME AND PRINCIPE
SUR SURINAME
SVK SLOVAK REPUBLIC
SVN SLOVENIA
SWE SWEDEN
SWZ SWAZILAND
SYC SEYCHELLES
SYR SYRIA
TCA TURKS AND CAICOS ISLANDS
TCD CHAD
TGO TOGO
THA THAILAND
TJK TAJIKISTAN
TKL TOKELAU
TKM TURKMENISTAN
TMP TIMOR-LESTE
TON TONGA
TTO TRINIDAD&TOBAGO
TUN TUNISIA
TUR TURKEY
TUV TUVALU
TWN TAIWAN
TZA TANZANIA
UGA UGANDA
UKR UKRAINE
URY URUGUAY
USA U.S.A.
UZB UZBEKISTAN
VCT ST.VINCENT&GRE
VEN VENEZUELA
VGB BRITISH VIRGIN ISLANDS
VIR VIRGIN ISLANDS (U.S.)
VNM VIET NAM
VUT VANUATU
WBG WEST BANK AND GAZA
WLF WALLIS AND FORTUNA
WSM WESTERN SAMOA
YEM YEMEN
YUG YUGOSLAVIA
ZAF SOUTH AFRICA
ZAR ZAIRE
ZMB ZAMBIA
ZWE ZIMBABWE
ZZZ OTHER TERRITORIES

APPENDIX 5.C

TABLES ROBUSTNESS CHECKS

Table 5.4b– Log current price GDP pc 1996, and foreign market access (NBER-UN)

Log GDP pc	(1)	(2)	(3)	(4)	(5)
Observations	95	58	95	27	95
Year	1996	1996	1996	1996	1996
LogFMA1994	.1210	-.2772	.0594	.0155	.0011
	(.0805)	(.1718)	(.0816)	(.0633)	(.0685)
Loghydroc pc	.0382***	.0413**	.0520***	.0039	.0493***
	(.0145)	(.0179)	(.0138)	(.0119)	(.0124)
Arable Land	-.0658	-.0547	-.0581	-.0596*	-.0510
	(.0459)	(.0695)	(.0422)	(.0341)	(.0420)
N minerals	.0038	.0077	.0043	.0020	.0047
	(.0101)	(.0112)	(.0103)	(.0094)	(.0076)
% Tropics	-.2506	-.4165***	-.4021*	.1532	.0778
	(.2006)	(.1460)	(.2027)	(.2655)	(.1494)
Malaria 1994	-1.145***	-1.412***	-.7723***	-2.182***	-1.056***
	(.2156)	(.2382)	(.2198)	(.2258)	(.1833)
Property rights	-.4466***	-	-	-	-.1776**
	(.0810)				(.0823)
Risk exprop	-	.1973***	-	-	-
		(.0470)			
Econ Free	-	-	-.8120***	-.7566***	-
			(.1450)	(.1695)	
Socialist	-.1242	-.0575	-.0313	(dropped)	-
	(.1774)	(.2332)	(.2060)		
War dummy	-.0377	-.0692	-.1220	(dropped)	-
	(.1585)	(.1614)	(.1742)		
Trust	-	-	-	.0167***	-
				(.0051)	
Gvt Effective	-	-	-	-	.5101***
					(.0862)
Estimation	OLS	OLS	OLS	OLS	OLS
R^2	.7833	.7152	.7982	.9321	.8433
F(.)	48.07	16.99	50.00	80.63	65.07
Prob>F	.0000	.0000	.0000	.0000	.0000

Constant not shown. Heteroskedasticity corrected White-robust standard errors in parentheses. *Statistically significant at the 90% level. **Statistically significant at the 95% level. ***Statistically significant at the 99% level.

Note: Updating the log GDP per capita from Penn World Table 6.1 to the new version 6.2 adds just one more observation into the operative sample. Given that the results are very similar in any case, the use of Penn World Table 6.1 GDP data is preferred for consistency.

Table 5.5b – Log of current price real GDP pc 1996 (PWT 6.1), Market Access (NBER-UN), and Governance Matters V

Log GDP pc	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Observations	95	95	95	95	95	93	93
Year	1996	1996	1996	1996	1996	1996	1996
LogTMA 1994	.0104	.0594	.0013	.0617	.0400	.0217	-.00003†
	(.0769)	(.0761)	(.0685)	(.0790)	(.0835)	(.0798)	(.0754)
Loghydroc pc	.0542***	.0463***	.0493***	.0534***	.0500***	.0461***	.0500***
	(.0136)	(.0135)	(.0124)	(.0141)	(.0147)	(.0130)	(.0142)
Arable Land	-.0683	-.0760	-.0510	-.0722	-.0734	-.0533	-.0384
	(.0431)	(.0441)	(.0420)	(.0418)	(.0447)	(.0417)	(.0400)
N minerals	.0033	.0070	.0047	.0024	.0014	.0082	.0036
	(.0090)	(.0089)	(.0076)	(.0100)	(.0098)	(.0082)	(.0081)
% Tropics	-.1644	-.1686	.0778	-.3200	-.2440	.0663	.1074
	(.1680)	(.1776)	(.1494)	(.1673)	(.1658)	(.1498)	(.1718)
Malaria 1994	-1.039***	-1.093***	-1.056***	-.7624***	-1.058***	-1.070***	-1.041***
	(.1903)	(.2016)	(.1833)	(.2171)	(.1966)	(.1872)	(.2403)
Property rights	-.3235***	-.3720***	-.1776**	-.2189**	-.3658***	-.2506***	-.1344
	(.0831)	(.0693)	(.0823)	(.0888)	(.0815)	(.0789)	(.0835)
Voice	.3351***	-	-	-	-	-	.0281
	(.0790)						(.1892)
Pol. stability	-	.2377***	-	-	-	-	.0781
		(.0747)					(.0832)
Gvt Effective	-	-	.5101***	-	-	-	.4775***
			(.0862)				(.1597)
Regulatory	-	-	-	.4965***	-	-	.0400
				(.1157)			(.1872)
Rule of Law	-	-	-	-	.2342**	-	-.0133
					(.0908)		(.1370)
Contrl corrup	-	-	-	-	-	.4074***	-.0068
						(.0762)	(.1404)
Estimation	OLS	OLS	OLS	OLS	OLS	OLS	OLS
R^2	.8104	.8033	.8433	.8132	.7972	.8246	.8509
F(.)	66.13	60.06	65.09	63.97	50.66	58.11	50.61
Prob>F	.0000	.0000	.0000	.0000	.0000	.0000	.0000

Constant not shown. Heteroskedasticity corrected White-robust standard errors in parentheses. *Statistically significant at the 90% level. **Statistically significant at the 95% level. ***Statistically significant at the 99% level. †5 significant ciphers instead of the regular 4 allowed.

REFERENCES

Secondary sources

Amjadi, Azita, and Alexander Yeats, 1995, 'Have Transport Costs Contributed to the Relative Decline of African Exports? Some Preliminary Empirical Evidence', World Bank Policy Research Working Paper (Washington, DC).

Blainey, Geoffrey, 1966, *The Tyranny of Distance: How Distance Shaped Australia's History*, Sun Books (Melboune, Australia).

Dixit, Avinash K., and Joseph E. Stiglitz, 1977, 'Monopolistic Competition and Optimum Product Diversity', *American Economic Review*, 67 (3):297-308.

Fujita, Masahisa, Paul Krugman, and Anthony J. Venables, 1999, *The Spatial Economy: Cities, Regions, and International Trade*, MIT Press (Cambridge, MA), chapter 14.

Gallup, John Luke, Jeffrey D. Sachs, and Andrew D. Mellinguer, 1999, 'Geography and Economic Development', *International Regional Science Review*, 22 (2):179-232.

Gallup, John Luke, and Jeffrey D. Sachs, 2001 January, 'The Economic Burden of Malaria', *American Journal of Tropical Medicine and Hygiene*, 64:85-96.

Head, Keith, and Thierry Mayer, 2002 January, 'Illusory Border Effects: Distance Mismeasurement Inflates Estimates of Home Bias in Trade', CEPII working paper 2002-01.

Knack, S., and P. Keefer, 1997, 'Does Social Capital Have an Economic Payoff?', *Quarterly Journal of Economics*, 112:1251-88.

Limao, Nuno, and Anthony J. Venables, 2001, 'Infrastructure, Geographical Disadvange, Transport Costs, and Trade', *The World Bank Economic Review* 15 (3): 451-479.

Redding, Stephen, and Anthony J. Venables, 2004, 'Economic Geography and International Inequality', *Journal of International Economics* 62: 53-82.

Temple, Jonathan, and Paul A. Johnson, 1998, 'Social Capability and Economic Growth', *The Quarterly Journal of Economics*, 113 (3) August:965-90.

Data sources

Adelman, Irma, and Cynthia Taft Morris, 1967, *Society, Politics, and Economic Development: A Quantitative Approach*, The Johns Hopkins Press (Baltimore).

Boston College, Faculty Microcomputer Resource Center (FMRC), 2000, Graduate Statistical Assistant Program, *Geography and Economic Development.* Access to *Geodata.*

Canadian Statistics Trade Data Archive.

CEPII – Centre d'Etudes Prospectives et d'Informations Internationales, Bases de Données, Distances, http://www.cepii.fr/francgraph/bdd/distances.htm

CEP-NBER World Trade Data, 2001. CEP internal data.

Clair, Godefroy, Guillaume Gaulier, Thierry Mayer, and Soledad Zignago, 2004, 'Notes on CEPII's distances measures', http://www.cepii.fr/distance/noticedist_en.pdf

European Values Study Foundation, The, and World Values Survey Association, 2006, *European and World Values Surveys Four-Wave Integrated Data File, 1981-2004, v.20060423*, Aggregate File Producers: ASEP/JDS, Madrid, Spain/Tilburg University, Tilburg, the Netherlands. Aggregate File Distributors: ASEP/JDS and ZA, Cologne, Germany.

Feenstra, Robert C., Robert E. Lipsey, Haiyan Deng, Alyson C. Ma, and Hengyong Mo, 2005 January, 'World Trade Flows: 1962-2000', NBER Working Paper Series 11040, http://www.nber.org/papers/w11040

Gallup, John Luke, and Jeffrey D. Sachs, with Andrew Mellinger, 1999 March, 'Geography and Economic Development', CID working paper 1, http://www2.cid.harvard.edu/cidwp/001.pdf

Heston, Alan, Robert Summers, and Bettina Aten, 2002, October, *Penn World Table Version 6.1*, Center for International Comparisons at the University of Pennsylvania (CICUP), http://pwt.econ.upenn.edu/php_site/pwt_index.php

——— , 2006, September, *Penn World Table Version 6.2*, Center for International Comparisons of Production, Income and Prices at the University of Pennsylvania.

Index of Economic Freedom, The Heritage Foundation and The Wall Street Journal, http://www.heritage.org/index/

International Association for Medical Assistance to Travelers, 2006 March 1, *World Malaria Risk Chart.*

Kaufmann, Daniel, Aart Kraay and Massimo Mastruzzi, 2006, 'Governance Matters V: Governance Indicators for 1996-2005', World Bank Policy Research September 2006.

NBER-UN Trade Data, 1962-2000, www.nber.org/data

Parker, Philip M., 1997, *National Cultures of the World: A Statistical Reference.* Cross-Cultural Statistical Encyclopedia of the World, vol. 4, Greenwood Press (Westport, CT).

Sachs, Jeffrey D. and Andrew M. Warner, 2005, 'Natural Resource Abundance and Economic Growth', NBER Working Paper 5398. Revised 1997, available at CID Harvard.

Stewart, Martin, 2001, 'The World Trade Database, 1970-97', CEP internal manuscript.

World Bank, 2005 July, 'World Bank List of Economies', http://siteresources.worldbank.org/DATASTATISTICS/Resources/CLASS.XLS

CONCLUSIONS

When I started this project I had many questions. Now I have many questions and some answers. The questions were revolving around the idea that the different ways people interact in different parts of the world must have influenced economic transactions. Civic virtues, mutual trust, voluntary work or spontaneous cooperation might arise spontaneously but constitute a flow and stock of a sort of abstract type of capital for the whole of the society, and contribute generating added value. So there must be a way to account for these differences, and to assign a value to the difference. There must be a way to measure the importance of social capital.

Later on I learnt that there was a recently coined term, social capital, which would refer precisely to this idea and there was a prolific literature on the subject, especially in political science and sociology. The new term had also had some impact in economics, a small stream quantifying the impact of social arrangements and, above all, trust, had on the economy. But my main question was how does this old idea with new terminology relate to the way different economies had evolved over time? Is social capital related and affecting the economy the same way in the past as it does now? If so, by how much? Are we talking about the same magnitude of relevance? In order to answer this question, I had first to be able to find a way to measure or capture social capital nowadays and in the past. And, if possible, in a way that would be directly comparable for both periods.

The Adelman and Morris (1967) work was illuminating. They had created an international Social Development Index (SDI) for the 1960s that was able to encapsulate the concept of the value of good social interactions and capture it indirectly at the same time. Not only this, but the later study of Temple and Johnson (1998) demonstrated that the Adelman and Morris' Index could have predicted future economic growth for a cross-section of countries better than any other contemporary attempt.

The statistical technique used to compute the index, the Principal Components Analysis (PCA) was exceptionally appropriate for extracting hidden dimensions out of a large amount of highly interrelated variables. This technique was out of use in mainstream Economics and unknown in Economic History, but is now being re-discovered and

widely used and accepted for the creation of indices (see Kauffman from the World Bank for the aggregation of governance related variables).

Reading all the works by Adelman and Morris I found that they invested a vast amount of time (more than 20 years of research) in compiling a massive homogeneous international database for the second half of the nineteenth century and beginning of the twentieth with many variables referring to society and economics. The purpose of this database was related to development studies; they grouped all the sample in different groups of countries and found differentiated general paths of development, some countries more based on agriculture, some more on technologies, and so on. I realised that this database had many more possibilities and could, indeed, be of great use for my social development index purposes.

Then it came the painstaking work of reconstruction and updating of the database. The choice of some of the variables to be included, excluded or updated deserved some thought. It is all explained in detail in chapter 2 (database). After some time, I had a completely new database, which would be ready to use as a basis for calculation of the new social index.

I believe the new database is a contribution to Economic History by its own right. This database is now digital, updated, and readily available for further studies. Overall, it includes 73 variables referring to aspects of society, politics and economics for 23 countries, for the period 1850-1914, some of them starting as early as 1830 (all listed in chapter 2). The new database has the potential to be used as a base for international comparative studies in Economic and Social History or related disciplines. It should be readily available upon request.

The next phase of the PhD was calculation of the Social Development Index (SDI). The selection of variables was done according to statistical adequacy criteria. 18 variables were included in the index (careful explanation of the variable selection process can be found in chapter 3). The new SDI was generated for 2 years: 1870 and 1980. And this is (imperfectly) comparable to the previously existing SDI for 1960. It is a pity that the sample of countries coinciding in both periods is small. So there is a lot more intense work to be done in that direction. However, this would take enormous amount of effort and imagination, since some of the variables cannot really be replicated again for new

countries because the raw basic information is based on the unreleased reports of country experts with no detailed criteria. Of course, I do encourage going along this avenue of widening the sample but I would suggest strict selection of feasible variables for future research.

Should I recommend going in any direction at all at this respect I would suggest widening the time dimension. Recall we now have the SDI for 1870, 1890 and 1960. These are good enough to show that it is possible to add a historical dimension to social capital measurement. However, there are three ways to go: In the first place, one could widen the time horizon going backwards. This is maybe the alternative that would take the least effort and be less time consuming because the database starts in 1850 and thus most of the variables are already available for that year. SDI for 1850 was not calculated in this thesis because some of the variables required lags and would have needed the data corresponding to 1830 to be able to carry it on. Exceptionally, some variables are available for 1830 but these are the exception and constructing a SDI for 1830 would certainly prove to be an extremely difficult task due to scarcity of reliable information. Mitchell (1998) points out that historical statistics start to be generally reliable from the second half of the nineteenth century onwards.

The second alternative would be to produce a SDI for intermediate time cuts, namely 1910 or alternatively 1914, 1930 or 1940 (options to be discussed). The first intermediate time cut is feasible and almost readily available, with the exception of variables that require forward moving information, because 1914 is the final year of the original Adelman and Morris (1988) database. A SDI for some arbitrary time cut between 1930 and 1940 is more difficult to construct because of lack of data. There is no clear date to assign to this proposed time cut, but I would do not encourage the choice of 1930 as cutting year because of the highly disruptive situation of the world economy (and society) during the Great Depression; in the same way the original database does not finish in 1910 but in 1914, just before World War I. If any, I would recommend some year that is already recognised in the literature as the end of the Great Depression or even later, but not beyond 1940. This would probably be 1939.

Finally, the third way to widen the time horizon of SDI would be to produce indices for the years after 1960 and all the way forward to present day. This implies some compilation of data, but the main advantage is that more recent times have statistics and

other information based on surveys readily available and generally easily accessible. The only question here would be whether it makes sense to generate an indirect general index when we have more precise concepts like mutual trust reported in rounds of international surveys such as the World Value Surveys. It does make sense, nevertheless, to construct such an index for the sake of comparison with the already existing years.

This is all as far as the proposed chronological extensions of the SDI are concerned. Other avenues for future research involving the enlargement of the sample of countries would be more difficult to execute, because the criteria to assign a specific value to qualitative variables was done on a comparative basis with respect to other countries and the decision processes are very difficult to recreate for many of the qualitative variables.

The second obvious contribution of the thesis for economic historians is the new index. The SDI is available for two time cuts, 1870 and 1890, and for 23 countries around the world. These new ranking figures are mainly thought for international comparisons, because the index is a ranking. At this respect, we can draw parallelisms between the SDI and the Human Development Index (HDI). The HDI is an index composed of several variables covering different aspects that attempts at giving an indication of the level of human development. The HDI has been a successful index, being widely used in development studies. In the same way, the SDI is an index composed of several variables that attempts at giving an indication of the level of social development, and may be useful in a broad range of disciplines within the social sciences, ranging from political science to social anthropology, as an indicator of the quality of society or social development.

The third contribution of the thesis is the knowledge derived from the empirical tests on the relationship between social development and economic development. This relationship has been shown to exist in earlier studies, but now has been confirmed to exist in the long run as well, so it becomes a stylised fact. Moreover, the associated coefficients reveal similar magnitudes for the nineteenth and twentieth centuries (around 50 to 60 percent of a standard deviation increase in accumulated long run growth for every 1 standard deviation increase in SDI for maximum correlation periods), which points at the existence of a stable positive relationship between social development and

income. More tests could be done with different benchmark years, but the results achieved up to now are robust enough to illustrate this point. Social capital and related concepts bring value to the economy.

There is one interesting finding that invites to further research, and this is the fact that the closer we move into the present the shortest is the lag with which SDI operates in the economy; it starts with 5-6 decades in the nineteenth century and seems to be decreasing over time. This finding seems to suggest that things move faster nowadays than in the past. The empirical results presented in chapter 4 go along this direction and are enough to make us think about this phenomenon, but are by no means sufficient to constitute a radical proof. Thus, it could be promising to investigate further in this direction.

Moving to the Economic Geography contributions, the forth contribution of the thesis is the new data on access to markets, starting in 1962, including maps and graphs. The Market Access indicator was recently designed was first presented by its developers for one year only, 1994 (Redding and Venables, 2004). In this thesis I calculated Market Access series for virtually all countries in the world starting in 1962, the initial year of the bilateral trade database. This can help studying the evolution of the index from the 1960 to nowadays, and it is just the start of an ongoing research project.

The fifth and final mayor contribution of the thesis are the conclusions drawn about the adequacy of institutions in countries that are distant to the main world markets. Some countries like New Zealand are far away from the main world markets and nevertheless have been able to achieve a healthy wealth status. Others like Zimbabwe are equally far away from the main world markets and remain poor. What is the difference between these two? Would this change should we find a way to bring them closer to markets? The exploration conducted in chapter 5 brings us to the institutions' debate. An institutional quality threshold is found, below which countries could not benefit from a favourable location. Then, the SDI helps testing the Market Access matters' hypothesis in a historical perspective. The result is that the findings about the existence of a minimum threshold of institutional quality are confirmed. The same institutional impediments that are being discussed nowadays in the institutional economics literature seem to have been present half a century ago. These findings suggest that there is a lot to be done in terms of policy recommendations to change institutions the way they are

operating in some countries because this could unleash the economic potential unrealised until now.

The SDI has facilitated the test of an economic geography theory for an alternative benchmark year. But this is just an example of the many more applications the index could have. The SDI can have many more applications into economic theory and other fields of the social sciences, apart from the purely descriptive mission of providing a world ranking of countries. Hopefully, future research will profit from it.

Despite the fact that this thesis does not enter into discussing the process of creation of social capital but only into the quantification and economic implications, policy recommendations regarding social capital would go along the way of encouraging measures that promote its emergence. In other words, measures encouraging a change of attitudes towards the other citizens; for instance, promoting new horizontal and voluntary organisations that would tighten the links within a community and between different groups of people.

Is economic history and the discipline of economics in general going to change its way of thinking and proceeding after this thesis? Certainly not, but it is true that this thesis is a marginal contribution towards the quantification of social capital and related variables, and a new idea in terms of looking at the social capital issue in historical perspective within economic history. This thesis is one more of the many works suggesting that the way people interact does matter and cannot be ignored any more in quantitative studies.

REFERENCES

Adelman, Irma, and Cynthia T. Morris, 1967, *Society, Politics, and Economic Development: A Quantitative Approach*, John Hopkins Press (Baltimore).

Kaufmann, Daniel, Aart Kraay and Pablo Zoido-Lobaton, 1999, *Aggregating Governance Indicators*, The World Bank (Washington, DC).

Mitchell, Brian Redman, 1998, *International Historical Statistics: Europe 1750-1993*, Fourth edition, Stockton Press and MacMillan Reference Ltd, (New York and Basingstoke respectively).

Redding, Stephen, and Anthony J. Venables, 2004, 'Economic Geography and International Inequality', *Journal of International Economics* 62: 53-82.

Temple, Jonathan, and Paul A. Johnson, 1998, 'Social Capability and Economic Growth', *The Quarterly Journal of Economics* 113 (3):965-990.